

2026-02

$\beta \ddot{y} \check{s} \pm \ddot{A} \pm \frac{1}{2} \pm \gg \acute{E} \ddot{A}^{10} - \hat{A} \ddot{A} \neg \tilde{A} \mu^1 \hat{A} \tilde{A} \ddot{A}^1 \hat{A}$
 $\beta \ddot{y} \cdot \gg \mu^0 \ddot{A} \acute{A} \zeta \frac{1}{2}^{10} - \hat{A} \pm^3 \zeta \acute{A} - \hat{A} \frac{1}{4} - \tilde{A} \acute{E} \pm \frac{1}{4}$
 $\beta \ddot{y} \acute{\prime} \mu \acute{\prime} \zeta \frac{1}{4} - \frac{1}{2} \acute{E} \frac{1}{2}^0 \pm^1 \frac{1}{4} \zeta \frac{1}{2} \ddot{A} - \gg \acute{E} \frac{1}{2} \frac{1}{4} \cdot$

$\beta \ddot{y} \acute{\prime} \neg^2 \pm \gg \zeta \acute{A}, \mu \acute{A} \uparrow \textcircled{R}$

$\beta \ddot{y} \textcircled{R} \mu \ddot{A} \pm \hat{A} \ddot{A} \acute{A} \zeta^{10} \hat{A} \hat{A} \pm \frac{1}{4} \frac{1}{4} \pm \tilde{A} \ddot{A} \cdot \frac{1}{2} \acute{\prime} \frac{1}{2} \neg \gg \acute{A} \tilde{A} \cdot \mu \acute{\prime} \zeta \frac{1}{4} - \frac{1}{2} \acute{E} \frac{1}{2}^0 \pm^1 \textcircled{R} \acute{A} \cdot \frac{1}{4} \pm \ddot{A} \zeta \zeta^{10} \zeta \frac{1}{2} \zeta$
 $\beta \ddot{y} \textcircled{R} \zeta \gg \textcircled{R} \check{Y}^{10} \zeta \frac{1}{2} \zeta \frac{1}{4}^{10} \hat{A} \frac{1}{2} \cdot \hat{A}^1 \tilde{A} \ddot{A} \cdot \frac{1}{4} \hat{A} \frac{1}{2}^0 \pm^1 \textcircled{R} \textcircled{R} \zeta^{-0} \cdot \tilde{A} \cdot \hat{A}, \pm \frac{1}{2} \mu \hat{A}^1 \tilde{A} \ddot{A} \textcircled{R} \frac{1}{4}^1 \zeta \cdot \mu \neg \hat{A} \zeta \gg^1 \hat{A}$

<http://hdl.handle.net/11728/13391>

Downloaded from HEPHAESTUS Repository, Neapolis University institutional repository



**ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ ΣΤΗΝ
ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ ΚΑΙ
ΧΡΗΜΑΤΟΟΙΚΟΝΟΜΙΚΗ ΤΕΧΝΟΛΟΓΙΑ**

Πανεπιστήμιο Νεάπολις Πάφος.

**«ΚΑΤΑΝΑΛΩΤΙΚΕΣ ΤΑΣΕΙΣ ΣΤΙΣ
ΗΛΕΚΤΡΟΝΙΚΕΣ ΑΓΟΡΕΣ ΜΕΣΩ ΑΝΑΛΥΣΗΣ
ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΜΟΝΤΕΛΩΝ ΜΗΧΑΝΙΚΗΣ
ΜΑΘΗΣΗΣ»**

Βάβαλου Περζή



**ΜΕΤΑΠΤΥΧΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΣΠΟΥΔΩΝ ΣΤΗΝ
ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ ΚΑΙ
ΧΡΗΜΑΤΟΟΙΚΟΝΟΜΙΚΗ ΤΕΧΝΟΛΟΓΙΑ**

**Διπλωματική Εργασία η οποία υποβλήθηκε προς απόκτηση
εξ αποστάσεως μεταπτυχιακού τίτλου σπουδών στην
ανάλυση δεδομένων και χρηματοοικονομική τεχνολογία στο
Πανεπιστήμιο Νεάπολις Πάφος.**

**«ΚΑΤΑΝΑΛΩΤΙΚΕΣ ΤΑΣΕΙΣ ΣΤΙΣ
ΗΛΕΚΤΡΟΝΙΚΕΣ ΑΓΟΡΕΣ ΜΕΣΩ ΑΝΑΛΥΣΗΣ
ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΜΟΝΤΕΛΩΝ ΜΗΧΑΝΙΚΗΣ
ΜΑΘΗΣΗΣ»**

Βάβαλου Περζή

ΦΕΒΡΟΥΑΡΙΟΣ, 2026

Πνευματικά δικαιώματα

Copyright © Βάβαλου Περζή, 2026

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Η έγκριση της παρούσας Διπλωματικής Εργασίας από το Πανεπιστήμιο Νεάπολις δεν υποδηλώνει απαραίτητως και αποδοχή των απόψεων του συγγραφέα εκ μέρους του Πανεπιστημίου.

Η ΥΠΕΥΘΥΝΗ ΔΗΛΩΣΗ

Η Βάβαλου Περζή, γνωρίζοντας τις συνέπειες της λογοκλοπής, δηλώνω υπεύθυνα ότι η παρούσα εργασία με τίτλο «Καταναλωτικές Τάσεις στις Ηλεκτρονικές Αγορές μέσω Ανάλυσης Δεδομένων και Μοντέλων Μηχανικής Μάθησης», αποτελεί προϊόν αυστηρά προσωπικής εργασίας και όλες οι πηγές που έχω χρησιμοποιήσει, έχουν δηλωθεί κατάλληλα στις βιβλιογραφικές παραπομπές και αναφορές. Τα σημεία όπου έχω χρησιμοποιήσει ιδέες, κείμενο ή/και πηγές άλλων συγγραφέων, αναφέρονται ευδιάκριτα στο κείμενο με την κατάλληλη παραπομπή και η σχετική αναφορά περιλαμβάνεται στο τμήμα των βιβλιογραφικών αναφορών με πλήρη περιγραφή.

Η Δηλούσα

Βάβαλου Περζή

Πίνακας περιεχομένων

.....	1
Περίληψη.....	9
Abstract	9
ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ	11
1.1 Στόχοι	11
1.2 Προσδιορισμός του Προβλήματος (Problem Statement).....	12
1.3 Ερευνητικά Ερωτήματα (Research Questions)	13
1.4 Σημασία και Προσδοκώμενη Συνεισφορά της Έρευνας.....	13
1.5 Διάρθρωση της Εργασίας	13
ΚΕΦΑΛΑΙΟ 2: ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ	14
2.1 Καταναλωτική συμπεριφορά στις ηλεκτρονικές αγορές	14
2.2 Ψηφιακές πλατφόρμες και επιρροή στο marketing.....	15
2.3 Συλλογή και είδη δεδομένων καταναλωτή.....	16
2.4 Ανάλυση δεδομένων και Μηχανική Μάθηση στο ηλεκτρονικό εμπόριο.....	17
2.5 Ηθικά ζητήματα και προστασία προσωπικών δεδομένων.....	19
2.6 Συμπεράσματα	20
ΚΕΦΑΛΑΙΟ 3: Μεθοδολογία Έρευνας και Ανάλυσης Δεδομένων (Research Methodology)	22
3.1 Περιγραφή και Τεκμηρίωση του Συνόλου Δεδομένων	22
3.1.1 Προέλευση και Χρονικό Πλαίσιο	23
3.1.2 Δομή και Χαρακτηριστικά Μεταβλητών	23
3.2 Προεπεξεργασία και Διασφάλιση Ποιότητας Δεδομένων (Data Wrangling).....	24
3.2.1 Διαχείριση Ελλειπόντων Τιμών.....	25
3.2.2 Διαχείριση Αρνητικών Τιμών και Επιστροφών	25
3.2.3 Αντιμετώπιση Ακραίων Τιμών (Outliers).....	25
3.3 Μετασχηματισμός Δεδομένων: Το Μοντέλο RFM.....	26
3.3.1 Τυποποίηση Δεδομένων (Z-Score Standardization).....	27
3.4 Επιλογή Εργαλείων Ανάλυσης	27
3.5 Διαδικασία Τμηματοποίησης: Αλγόριθμος K-Means	28
3.6 Έλεγχοι Αξιοπιστίας και Εγκυρότητας	30
ΚΕΦΑΛΑΙΟ 4: Ανάλυση Δεδομένων και Αποτελέσματα 4.1 Εισαγωγή	31
4.2 Περιγραφική Ανάλυση Δεδομένων (Descriptive Analytics)	31
4.2.1 Γεωγραφική Κατανομή και Διασυννοριακό Εμπόριο.....	31
4.2.2 Χρονική Εξέλιξη και Φαινόμενα Εποχικότητας.....	32
4.3 Ανάλυση Συμπεριφοράς Πελατών (RFM Analysis)	33

4.3.1 Ανάλυση Προσφατότητας (Recency).....	33
4.3.3 Ανάλυση Χρηματικής Αξίας (Monetary)	34
4.4 Αποτελέσματα Τμηματοποίησης (Clustering Analysis).....	35
4.4.1 Ομάδα 1: «VIP / Πολύτιμοι Πελάτες» (The VIPs)	35
4.4.2 Ομάδα 2: «Πιστοί / Ανερχόμενοι» (Loyal / Potential)	36
4.4.3 Ομάδα 3: «Τυπικοί / Περιστασιακοί» (Occasional / Mass Market).....	36
4.5 Συζήτηση και Διοικητικές Προεκτάσεις (Managerial Implications)	38
4.6 Συγκριτική Αξιολόγηση και Επιστημονική Τεκμηρίωση Ευρημάτων	39
ΚΕΦΑΛΑΙΟ 5: ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΜΕΛΛΟΝΤΙΚΕΣ ΕΠΕΚΤΑΣΕΙΣ	41
5.1 Συμπεράσματα	41
5.2 Μελλοντικές Επεκτάσεις.....	42
Βιβλιογραφία.....	44

Πίνακας: Συντομογραφίες και Ακρωνύμια

Ακρωνύμιο	Πλήρης Όρος (Αγγλικά)	Όρος στα Ελληνικά
RFM	Recency, Frequency, Monetary	Προσφατότητα, Συχνότητα, Χρηματική Αξία
CLV	Customer Lifetime Value	Δια Βίου Αξία Πελάτη
GDPR	General Data Protection Regulation	Γενικός Κανονισμός Προστασίας Δεδομένων
ROI	Return on Investment	Απόδοση Επένδυσης
CAC	Customer Acquisition Cost	Κόστος Απόκτησης Πελάτη
UCI	University of California, Irvine	Πανεπιστήμιο της Καλιφόρνια (Πηγή Dataset)
K-Means	K-Means Clustering Algorithm	Αλγόριθμος Ομαδοποίησης K-Means

Καταναλωτικές Τάσεις στις Ηλεκτρονικές Αγορές μέσω Ανάλυσης Δεδομένων και Μοντέλων Μηχανικής Μάθησης

Όνοματεπώνυμο Φοιτητή: Βάβαλου Περζή

Τίτλος Διπλωματικής Εργασίας: ΚΑΤΑΝΑΛΩΤΙΚΕΣ ΤΑΣΕΙΣ ΣΤΙΣ
ΗΛΕΚΤΡΟΝΙΚΕΣ ΑΓΟΡΕΣ ΜΕΣΩ ΑΝΑΛΥΣΗΣ ΔΕΔΟΜΕΝΩΝ ΚΑΙ
ΜΟΝΤΕΛΩΝ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ

Η παρούσα Διπλωματική Εργασία εκπονήθηκε στο πλαίσιο των σπουδών για την απόκτηση εξ αποστάσεως μεταπτυχιακού τίτλου στο Πανεπιστήμιο Νεάπολις και εγκρίθηκε στις από τα μέλη της Εξεταστικής Επιτροπής.

Εξεταστική Επιτροπή:

Πρώτος επιβλέπων: Κωνσταντίνος Παναγιωτάκης, Καθηγητής,
Τμήμα ΔΕΤ, ΕΛΜΕΠΑ

Μέλος Εξεταστικής Επιτροπής: Χρήστος Λεμονάκης,
Αναπληρωτής Καθηγητής, Τμήμα ΔΕΤ, ΕΛΜΕΠΑ

Μέλος Εξεταστικής Επιτροπής: Γεώργιος Μαστοράκης,
Αναπληρωτής Καθηγητής, Τμήμα ΔΕΤ, ΕΛΜΕΠΑ

Ευχαριστίες

Θα ήθελα να εκφράσω τις θερμές μου ευχαριστίες στον καθηγητή μου Κωνσταντίνο Παναγιωτάκη, για την αμέριστη υποστήριξη, καθοδήγηση και τις πολύτιμες συμβουλές του καθ' όλη τη διάρκεια αυτής της διπλωματικής εργασίας. Η εμπειρία που απέκτησα υπό την καθοδήγησή του ήταν καθοριστική για την ανάπτυξη των γνώσεων μου και την εμβάθυνση στο αντικείμενο της έρευνας. Η υπομονή, το ενδιαφέρον και η αφοσίωσή του στην εκπαιδευτική διαδικασία, καθώς και οι ιδέες και οι προτάσεις του, υπήρξαν καθοριστικά για την ολοκλήρωση αυτού του έργου.

Τον ευχαριστώ για την εμπιστοσύνη που μου έδειξε και για τη συνεχιζόμενη ενθάρρυνση που μου προσέφερε. Η στήριξή του υπήρξε ανεκτίμητη και η καθοδήγησή του αποτέλεσε πηγή έμπνευσης και δημιουργικότητας.

Περίληψη

Η παρούσα διπλωματική εργασία διερευνά τις καταναλωτικές τάσεις στο σύγχρονο ψηφιακό περιβάλλον, εστιάζοντας στη μετατροπή των ακατέργαστων δεδομένων συναλλαγών σε στρατηγική γνώση. Μέσω της ανάλυσης του διεθνώς αναγνωρισμένου συνόλου δεδομένων "Online Retail Dataset" (UCI Machine Learning Repository), η έρευνα επιχειρεί τη δομημένη τμηματοποίηση της πελατειακής βάσης. Η μεθοδολογική προσέγγιση συνδυάζει το μοντέλο RFM (Recency, Frequency, Monetary) για την ποσοτικοποίηση της αγοραστικής συμπεριφοράς, με προηγμένες τεχνικές στατιστικής τυποποίησης (Z-Scores) και τον αλγόριθμο μηχανικής μάθησης K-Means Clustering. Η οπτικοποίηση και ανάλυση των δεδομένων μέσω του εργαλείου Tableau ανέδειξε τρεις (3) διακριτές στρατηγικές συστάδες πελατών (VIP, Πιστοί, Περιστασιακοί). Τα ευρήματα επιβεβαιώνουν την ισχύ της Αρχής του Pareto, αναδεικνύοντας μια κρίσιμη μειοψηφία πελατών που παράγει το μεγαλύτερο ποσοστό της κερδοφορίας. Η εργασία καταλήγει σε εξειδικευμένες διοικητικές προτάσεις, όπως η αυτοματοποιημένη διακράτηση (Retention) και η μεγιστοποίηση της Δια Βίου Αξίας του Πελάτη (CLV). Συνολικά, καταδεικνύεται ότι η σύγκλιση της Επιστήμης των Δεδομένων με το Στρατηγικό Μάρκετινγκ αποτελεί τον ακρογωνιαίο λίθο για τη λήψη τεκμηριωμένων επιχειρησιακών αποφάσεων στην εποχή του Digital Transformation..

Λέξεις κλειδιά: E-commerce, Consumer Behavior, RFM Model, K-Means Clustering, Z-Scores, Tableau, Marketing Strategy, Customer Lifetime Value (CLV), Digital Transformation.

Abstract

This Master's thesis investigates evolving consumer trends in the e-commerce landscape by transforming raw transactional data into actionable strategic insights. Utilizing the internationally recognized "**Online Retail Dataset**" (**UCI Machine Learning Repository**), the study implements a structured customer segmentation framework. The methodology integrates the **RFM (Recency, Frequency, Monetary)** model to quantify purchasing behavior, alongside advanced statistical normalization techniques (**Z-Scores**) and the **K-Means Clustering** machine learning algorithm. Data visualization and analysis conducted via **Tableau** revealed three (3) distinct strategic customer segments (**VIP, Loyal, and Occasional**). The findings validate **Pareto's Principle**, highlighting a critical minority of customers that generates the vast majority of total revenue. The thesis concludes with targeted managerial implications, including automated retention strategies and the optimization of **Customer Lifetime Value (CLV)**. Ultimately, the study demonstrates that the synergy between Data Science and Strategic Marketing is a fundamental pillar for evidence-based decision-making in the era of Digital Transformation.

Keywords: E-commerce, Consumer Behavior, RFM Model, K-Means Clustering, Z-Scores, Tableau, Marketing Strategies

Αφιερώνεται

«Στην οικογένειά μου και τους και τους φίλους μου, των οποίων η αγάπη, η υποστήριξη και η αφοσίωση υπήρξαν αναπόσπαστο μέρος της πορείας μου. Στην οικογένειά μου, για την αδιάκοπη ενθάρρυνση και τις θυσίες τους, που αποτέλεσαν θεμέλιο της προσωπικής και ακαδημαϊκής μου εξέλιξης. Στους φίλους μου για τη συνεχιζόμενη πίστη τους σε μένα και πηγή δύναμης που αποτέλεσαν σε κάθε βήμα αυτής της διαδρομής. Η στήριξή του ήταν και είναι πολύτιμη και τους οφείλω το γεγονός ότι αυτή η εργασία ολοκληρώθηκε με επιτυχία.»

ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ

1.1 Στόχοι

Το ηλεκτρονικό εμπόριο αποτελεί μία από τις πλέον σημαντικές και ανατρεπτικές εξελίξεις της σύγχρονης οικονομίας, μετασχηματίζοντας τον τρόπο με τον οποίο οι καταναλωτές αναζητούν, αξιολογούν και αποκτούν προϊόντα και υπηρεσίες (Shrivastava et al., 2023). Η ραγδαία τεχνολογική πρόοδος, η διάδοση των κινητών συσκευών και η διεξόδωση του διαδικτύου σε όλες τις πτυχές της καθημερινότητας έχουν δημιουργήσει ένα περιβάλλον όπου οι αγοραστικές αποφάσεις λαμβάνονται κυρίως μέσα από ψηφιακά κανάλια (Singh, 2024).

Στο πλαίσιο αυτό, η κατανόηση της καταναλωτικής συμπεριφοράς αποκτά στρατηγική σημασία για τις επιχειρήσεις, καθώς η δυνατότητα να εντοπίζουν και να ερμηνεύουν τις τάσεις της αγοράς συνδέεται άμεσα με την ανταγωνιστικότητά τους (Hafez, 2024). Οι επιχειρήσεις δεν περιορίζονται πλέον στην παραδοσιακή καταγραφή πωλήσεων· επιδιώκουν τη σε βάθος ανάλυση των προτύπων αγοραστικής συμπεριφοράς, των προτιμήσεων των χρηστών και των παραγόντων που επηρεάζουν τη διαδικασία αγοράς (Chen et al., 2022).

Η αξιοποίηση της ανάλυσης δεδομένων (data analytics) στις ηλεκτρονικές αγορές επιτρέπει την εξαγωγή χρήσιμων συμπερασμάτων από μεγάλους όγκους πληροφοριών, όπως το ιστορικό συναλλαγών, η συχνότητα αγορών, οι προτιμήσεις προϊόντων και τα δημογραφικά χαρακτηριστικά (Wamba et al., 2023). Μέσω τεχνικών όπως η εξερευνητική ανάλυση δεδομένων (Exploratory Data Analysis – EDA) και η ομαδοποίηση (clustering), είναι δυνατόν να αναδειχθούν μοτίβα συμπεριφοράς και να προσδιοριστούν ομάδες καταναλωτών με κοινά χαρακτηριστικά (Levine et al., 2023).

Παρά την πληθώρα διαθέσιμων δεδομένων, πολλές επιχειρήσεις αντιμετωπίζουν δυσκολίες στην ουσιαστική αξιοποίησή τους. Συχνά περιορίζονται σε βασικούς δείκτες απόδοσης, όπως ο συνολικός αριθμός πωλήσεων ή το ποσοστό μετατροπών (conversion rate), χωρίς να εμβαθύνουν στις υποκείμενες τάσεις που διαμορφώνουν αυτές τις μεταβολές. Η ποικιλομορφία των πηγών δεδομένων —κοινωνικά δίκτυα, πλατφόρμες ηλεκτρονικού εμπορίου, συστήματα CRM— καθιστά την ανάλυση ακόμη πιο περίπλοκη και απαιτεί εξειδικευμένες μεθόδους και εργαλεία (Shmueli et al., 2017).

Η παρούσα εργασία επιδιώκει να αναδείξει τη σημασία της ανάλυσης δεδομένων ως εργαλείου κατανόησης της καταναλωτικής συμπεριφοράς, εστιάζοντας όχι στην πρόβλεψη μελλοντικών συμπεριφορών, αλλά στην ερμηνεία των υφιστάμενων τάσεων. Μέσα από την εφαρμογή των τεχνικών EDA και clustering επιχειρείται η κατανόηση των προτύπων που χαρακτηρίζουν διαφορετικές κατηγορίες καταναλωτών στις ηλεκτρονικές αγορές, παρέχοντας πρακτικές γνώσεις για στρατηγικές μάρκετινγκ (Zenodo, 2024).

Η δομή της εργασίας ακολουθεί μια λογική πορεία: αρχικά παρουσιάζεται το θεωρητικό πλαίσιο, ακολουθεί η περιγραφή της μεθοδολογίας και των εργαλείων, έπειτα η παρουσίαση των αποτελεσμάτων με την ερμηνεία τους, ενώ στο τέλος

παρατίθενται τα συμπεράσματα, οι περιορισμοί και οι προτάσεις για μελλοντική έρευνα.

Η κύρια επιδίωξη της παρούσας διπλωματικής εργασίας συνίσταται στην εις βάθος κατανόηση των καταναλωτικών προτύπων στο περιβάλλον του ηλεκτρονικού εμπορίου, μέσω της συστηματικής ανάλυσης δεδομένων συναλλαγών. Η ραγδαία ανάπτυξη των ψηφιακών πλατφορμών και η συνεχής αύξηση του όγκου των ηλεκτρονικών αγορών καθιστούν την αποκωδικοποίηση της καταναλωτικής συμπεριφοράς κρίσιμη παράμετρο για τη λήψη τεκμηριωμένων και στρατηγικών επιχειρησιακών αποφάσεων (Kotler et al., 2020; Laudon & Traver, 2022).

Στο πλαίσιο αυτό, η έρευνα επικεντρώνεται πρωτίστως στην ανάλυση της αγοραστικής συμπεριφοράς μέσω της αξιολόγησης της πελατειακής αξίας, αξιοποιώντας αποκλειστικά δεδομένα συναλλαγών βάσει των δεικτών Προσφατότητας (Recency), Συχνότητας (Frequency) και Χρηματικής Αξίας (Monetary) (Chaffey, 2021). Παράλληλα, επιδιώκεται η κατάτμηση της πελατειακής βάσης σε ομοιογενείς ομάδες (clustering) με την εφαρμογή του αλγορίθμου K-Means και τεχνικών στατιστικής τυποποίησης (Z-Scores), με σκοπό την ανάδειξη των διαφοροποιήσεων που παρουσιάζουν οι καταναλωτές στη συμπεριφορά τους (Han et al., 2022; Tan et al., 2019).

Επιπρόσθετα, βασικό στόχο αποτελεί η αξιοποίηση των ερευνητικών ευρημάτων για τη διαμόρφωση και ενίσχυση στοχευμένων στρατηγικών μάρκετινγκ, οι οποίες θα εστιάζουν στην αύξηση της διατήρησης πελατών (customer retention) και στη βελτιστοποίηση της προσέγγισης διαφορετικών τμημάτων της αγοράς (Wedel & Kannan, 2016). Η ανάλυση προσανατολίζεται στην ερμηνεία των υφιστάμενων καταναλωτικών τάσεων μέσω της χρήσης διαδραστικών εργαλείων οπτικοποίησης, παρέχοντας τεκμηριωμένη πληροφόρηση που δύναται να υποστηρίξει τη διοικητική λήψη αποφάσεων (Laudon & Traver, 2022). Η επίτευξη των ανωτέρω στόχων υποστηρίζεται μεθοδολογικά μέσω της συλλογής και επεξεργασίας μεγάλου όγκου δευτερογενών δεδομένων, της εφαρμογής της Εξερευνητικής Ανάλυσης Δεδομένων (EDA) και της οπτικοποίησης των αποτελεσμάτων μέσω του εργαλείου Tableau (Han et al., 2022).

1.2 Προσδιορισμός του Προβλήματος (Problem Statement)

Παρά την πληθώρα διαθέσιμων δεδομένων, πολλές επιχειρήσεις αντιμετωπίζουν το παράδοξο της «αφθονίας δεδομένων αλλά έλλειψης γνώσης» (data rich, insight poor). Συχνά περιορίζονται σε βασικούς δείκτες απόδοσης, όπως ο συνολικός αριθμός πωλήσεων, χωρίς να εμβαθύνουν στις υποκείμενες τάσεις που διαμορφώνουν αυτές τις μεταβολές. Η πολυπλοκότητα των σύγχρονων ψηφιακών συναλλαγών απαιτεί την υιοθέτηση εξειδικευμένων μεθόδων και αλγορίθμων που μπορούν να μετατρέψουν τον θόρυβο των δεδομένων σε αξιοποιήσιμη στρατηγική πληροφόρηση (Shmueli et al., 2017). Η παρούσα εργασία επιχειρεί να γεφυρώσει αυτό το χάσμα, χρησιμοποιώντας το μοντέλο RFM και τεχνικές μηχανικής μάθησης για την αποκάλυψη της κρυμμένης αξίας της πελατειακής βάσης.

1.3 Ερευνητικά Ερωτήματα (Research Questions)

Προκειμένου να επιτευχθεί ο κεντρικός σκοπός της μελέτης, η έρευνα καλείται να απαντήσει στα ακόλουθα ερωτήματα:

- **RQ1:** Πώς μπορεί η εφαρμογή του μοντέλου RFM (Recency, Frequency, Monetary) να ποσοτικοποιήσει αποτελεσματικά την ετερογένεια της αγοραστικής συμπεριφοράς;
- **RQ2:** Ποιος είναι ο βέλτιστος αριθμός συστάδων (clusters) που εξασφαλίζει τη βέλτιστη ισορροπία μεταξύ στατιστικής ακρίβειας και επιχειρηματικής ερμηνευτικότητας;
- **RQ3:** Ποιες στοχευμένες στρατηγικές μάρκετινγκ μπορούν να αναπτυχθούν για κάθε διακριτή ομάδα πελατών με σκοπό τη μεγιστοποίηση της διακράτησης (retention) και της κερδοφορίας;

1.4 Σημασία και Προσδοκώμενη Συνεισφορά της Έρευνας

Η συνεισφορά της παρούσας μελέτης είναι διττή: ακαδημαϊκή και πρακτική. Σε ακαδημαϊκό επίπεδο, ενισχύει τη βιβλιογραφία σχετικά με τη χρήση του K-Means Clustering σε δεδομένα λιανικής, υπογραμμίζοντας την κρισιμότητα της στατιστικής τυποποίησης (Z-Scores) για την αξιοπιστία της ομαδοποίησης (Laudon & Traver, 2022).

Σε πρακτικό επίπεδο, η έρευνα παρέχει ένα εφαρμόσιμο πλαίσιο για τις επιχειρήσεις ώστε να μεταβούν από το μαζικό μάρκετινγκ στο μάρκετινγκ ακριβείας (precision marketing). Μέσω της οπτικοποίησης των αποτελεσμάτων στο Tableau, η εργασία προσφέρει εργαλεία λήψης αποφάσεων που επιτρέπουν τη στρατηγική εστίαση στους πελάτες υψηλής αξίας (VIP), βελτιστοποιώντας την κατανομή των πόρων και ενισχύοντας τη Δια Βίου Αξία του Πελάτη (CLV) (Han et al., 2022; Wedel & Kannan, 2016).

1.5 Διάρθρωση της Εργασίας

Η εργασία οργανώνεται σε πέντε διακριτά κεφάλαια ως εξής:

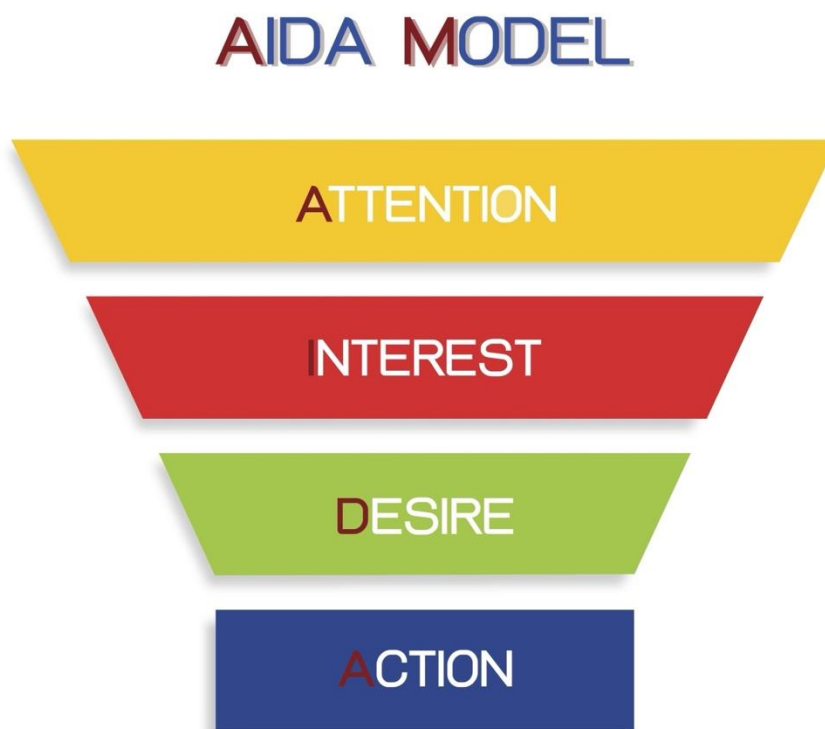
- **Κεφάλαιο 1:** Εισαγωγή, σκοπός, ερευνητικά ερωτήματα και συνεισφορά της μελέτης.
- **Κεφάλαιο 2:** Βιβλιογραφική επισκόπηση του ηλεκτρονικού εμπορίου, της καταναλωτικής συμπεριφοράς και των μεθοδολογιών RFM και Clustering.
- **Κεφάλαιο 3:** Παρουσίαση της μεθοδολογίας, της διαδικασίας καθαρισμού των δεδομένων και των εργαλείων ανάλυσης.
- **Κεφάλαιο 4:** Ανάλυση αποτελεσμάτων, οπτικοποίηση των συστάδων και συγκριτική αξιολόγηση με τη διεθνή βιβλιογραφία.
- **Κεφάλαιο 5:** Συνοπτικά συμπεράσματα, περιορισμοί της έρευνας και προτάσεις για μελλοντική επιστημονική διερεύνηση.

ΚΕΦΑΛΑΙΟ 2: ΘΕΩΡΗΤΙΚΟ ΥΠΟΒΑΘΡΟ

2.1 Καταναλωτική συμπεριφορά στις ηλεκτρονικές αγορές

Η καταναλωτική συμπεριφορά συνιστά ένα πολυδιάστατο πεδίο μελέτης που αφορά τις διαδικασίες μέσω των οποίων τα άτομα επιλέγουν, αγοράζουν, χρησιμοποιούν και αξιολογούν προϊόντα ή υπηρεσίες, με απώτερο σκοπό την ικανοποίηση των αναγκών και των επιθυμιών τους (Solomon, 2020). Στο πλαίσιο του ηλεκτρονικού εμπορίου, η συμπεριφορά αυτή αποκτά νέες διαστάσεις και πολυπλοκότητα, καθώς οι καταναλωτές αλληλεπιδρούν εντός ενός ψηφιακού περιβάλλοντος που χαρακτηρίζεται από υπερπροσφορά πληροφοριών, ευρύτατη ποικιλία επιλογών και εντατικό ανταγωνισμό (Laudon & Traver, 2022). Η μετάβαση από τις παραδοσιακές στις ηλεκτρονικές αγορές οφείλεται πρωτίστως στην ευκολία πρόσβασης, στη δυνατότητα άμεσης σύγκρισης τιμών και στη διαφάνεια που προσφέρουν οι διαδικτυακές πλατφόρμες (Kotler & Keller, 2023). Επιπροσθέτως, η δυνατότητα των χρηστών να έχουν πρόσβαση σε αξιολογήσεις άλλων καταναλωτών ή να λαμβάνουν προσωποποιημένες προτάσεις από αλγόριθμους συστάσεων, επηρεάζει καθοριστικά τη διαδικασία λήψης αποφάσεων (Chaffey & Ellis-Chadwick, 2022).

Η μελέτη της ηλεκτρονικής καταναλωτικής συμπεριφοράς βασίζεται σε θεμελιώδη θεωρητικά μοντέλα που ερμηνεύουν την αγοραστική απόφαση. Χαρακτηριστικό παράδειγμα αποτελεί το μοντέλο AIDA, το οποίο περιγράφει την κλιμακωτή πορεία του καταναλωτή μέσα από τα στάδια της προσοχής (Attention), του ενδιαφέροντος (Interest), της επιθυμίας (Desire) και τελικά της δράσης (Action) (Kotler & Keller, 2023). Παράλληλα, η Θεωρία της Προγραμματισμένης Συμπεριφοράς (TPB) του Ajzen (1991) υποστηρίζει ότι οι στάσεις, οι κοινωνικοί κανόνες και η αντιλαμβανόμενη ευκολία ελέγχου επιδρούν στην πρόθεση αγοράς. Στο πεδίο της τεχνολογικής αποδοχής, το μοντέλο UTAUT (Unified Theory of Acceptance and Use of Technology) ερμηνεύει την υιοθέτηση νέων τεχνολογιών βάσει παραγόντων όπως η προσδοκώμενη απόδοση και η κοινωνική επιρροή (Venkatesh et al., 2003).



Εικόνα 2.1: Το μοντέλο AIDA και τα στάδια της καταναλωτικής διαδρομής (Purchase Funnel). Πηγή: Προσαρμογή από Kotler & Keller (2023).

Στο περιβάλλον του ηλεκτρονικού εμπορίου, κρίσιμοι παράγοντες που διαμορφώνουν τη συμπεριφορά είναι η εμπιστοσύνη στην πλατφόρμα, η αντίληψη ασφάλειας των συναλλαγών και η ποιότητα της εμπειρίας χρήστη (UX) (Gefen et al., 2003). Στοιχεία όπως η ευκολία πλοήγησης, η ταχύτητα φόρτωσης της ιστοσελίδας και η διαφάνεια στην πολιτική επιστροφών συμβάλλουν θετικά στην ενίσχυση της πρόθεσης αγοράς (Laudon & Traver, 2022). Για παράδειγμα, ένας χρήστης που επιλέγει ένα ηλεκτρονικό κατάστημα ένδυσης επηρεάζεται συνδυαστικά από ψυχολογικούς και κοινωνικούς παράγοντες, όπως οι κριτικές άλλων πελατών, αλλά και από τεχνολογικούς παράγοντες, όπως η ασφάλεια των πληρωμών και η αξιοπιστία της διανομής.

2.2 Ψηφιακές πλατφόρμες και επιρροή στο marketing

Οι ψηφιακές πλατφόρμες, συμπεριλαμβανομένων των μέσων κοινωνικής δικτύωσης, των ηλεκτρονικών καταστημάτων και των εφαρμογών κινητών, έχουν αναμορφώσει ριζικά το τοπίο του μάρκετινγκ. Η παραδοσιακή, μονόδρομη επικοινωνία έχει πλέον εξελιχθεί σε μια διαδραστική και προσωποποιημένη εμπειρία μεταξύ επιχείρησης και καταναλωτή (Kaplan & Haenlein, 2020). Η καταλυτική επιρροή των ψηφιακών πλατφορμών έγκειται στην ικανότητά τους να συγκεντρώνουν και να αναλύουν μεγάλα σύνολα δεδομένων (Big Data), επιτρέποντας την υλοποίηση στοχευμένων εκστρατειών και την παροχή εξατομικευμένων εμπειριών (Chaffey & Ellis-Chadwick, 2022). Οι αλγόριθμοι που αξιοποιούνται από πλατφόρμες όπως η Meta, η Google και η Amazon επιτρέπουν την προβολή διαφημιστικού περιεχομένου σε

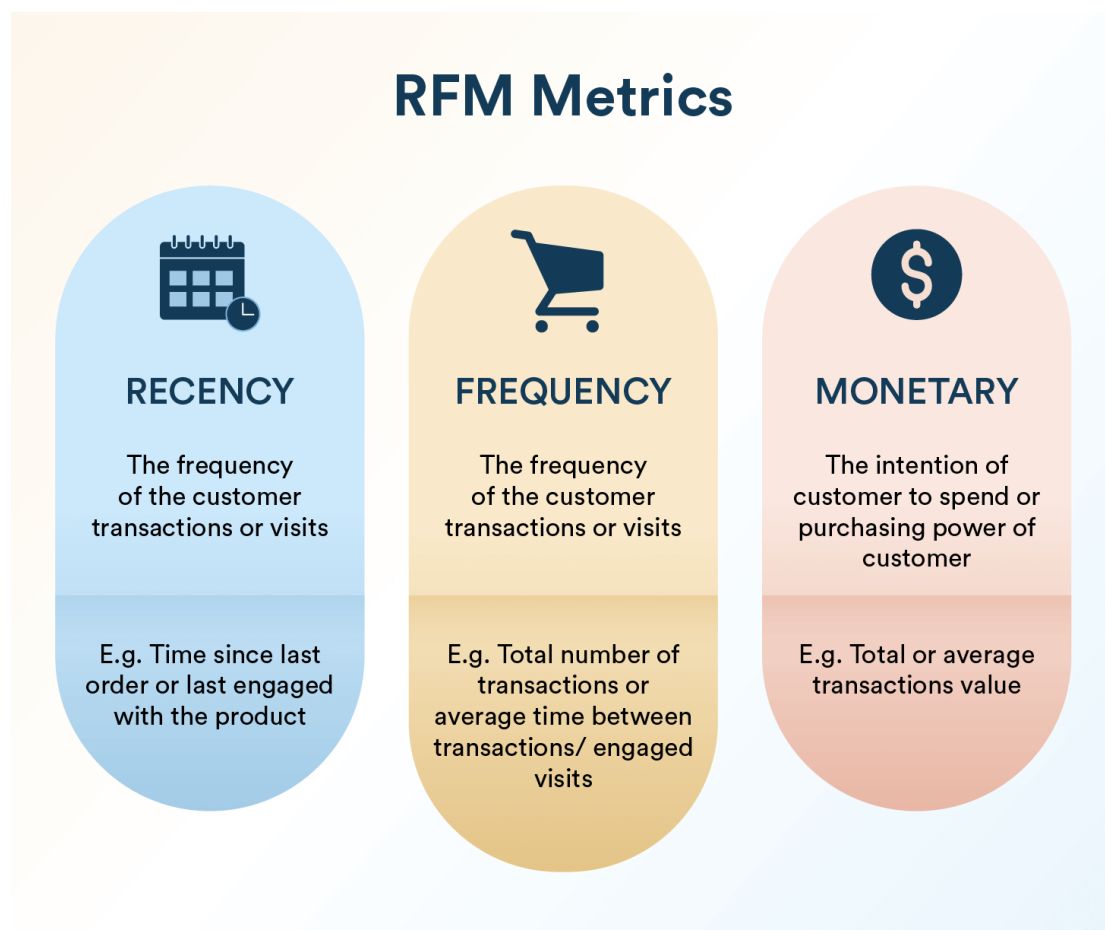
συγκεκριμένα δημογραφικά και συμπεριφορικά προφίλ, αυξάνοντας σημαντικά τα ποσοστά μετατροπής (conversion rates).

Επιπλέον, οι πλατφόρμες αυτές προάγουν την αμφίδρομη επικοινωνία, ενισχύοντας την αλληλεπίδραση και οικοδομώντας σχέσεις εμπιστοσύνης (Mangold & Faulds, 2009). Μέσω των σχολίων, των αξιολογήσεων και της κοινοποίησης περιεχομένου, οι καταναλωτές μετατρέπονται σε ενεργούς συνδιαμορφωτές της εταιρικής εικόνας. Σε αυτό το πλαίσιο, η εξατομίκευση περιεχομένου (personalization) αναδεικνύεται σε καθοριστικό παράγοντα επιτυχίας. Επιχειρήσεις που εφαρμόζουν στρατηγικές βασισμένες στην ανάλυση της συμπεριφοράς των χρηστών επιτυγχάνουν υψηλότερα επίπεδα δέσμευσης (engagement) και πιστότητας (loyalty) (Lemon & Verhoef, 2016). Χαρακτηριστικό παράδειγμα αποτελεί η Netflix, η οποία χρησιμοποιεί προηγμένα μοντέλα μηχανικής μάθησης για να προτείνει περιεχόμενο προσαρμοσμένο στις ατομικές προτιμήσεις, αυξάνοντας τον χρόνο παραμονής και μειώνοντας τα ποσοστά αποχώρησης (Gomez-Uribe & Hunt, 2016). Παράλληλα, η άνοδος του social commerce, δηλαδή της εμπορικής δραστηριότητας εντός κοινωνικών δικτύων, συνδυάζει την ψυχαγωγία με την κατανάλωση, δημιουργώντας ένα νέο περιβάλλον πωλήσεων που βασίζεται στην αυθεντικότητα και την κοινωνική επιρροή (Appel et al., 2020).

2.3 Συλλογή και είδη δεδομένων καταναλωτή

Η συλλογή και ανάλυση δεδομένων καταναλωτή αποτελεί τον πυρήνα του σύγχρονου ηλεκτρονικού εμπορίου, προσφέροντας στις επιχειρήσεις τη δυνατότητα να κατανοήσουν βαθύτερα το κοινό τους, να προβλέψουν τάσεις και να σχεδιάσουν αποτελεσματικότερες στρατηγικές (Wedel & Kannan, 2016). Τα δεδομένα που συλλέγονται κατηγοριοποιούνται σε τέσσερις βασικές ομάδες. Αρχικά, τα δεδομένα προτιμήσεων περιλαμβάνουν τις επιλογές και αλληλεπιδράσεις του χρήστη, όπως αναζητήσεις και κλικ, τροφοδοτώντας συστήματα συστάσεων (Aggarwal, 2016). Ακολούθως, τα δεδομένα αγοραστικής συμπεριφοράς αφορούν το ιστορικό αγορών, τη συχνότητα και τη δαπάνη, στοιχεία που είναι κρίσιμα για την πρόβλεψη της ζήτησης (Kotler & Keller, 2023). Τρίτον, τα δημογραφικά δεδομένα (ηλικία, τοποθεσία κ.ά.) αξιοποιούνται για την τμηματοποίηση της αγοράς (Laudon & Traver, 2022), ενώ τέλος, τα συμπεριφορικά και συναισθηματικά δεδομένα εξετάζουν την πλοήγηση και τη συναισθηματική ανταπόκριση (Lemon & Verhoef, 2016).

Ιδιαίτερη θεωρητική βαρύτητα για την παρούσα μελέτη κατέχουν τα δεδομένα αγοραστικής συμπεριφοράς και συγκεκριμένα το μοντέλο RFM (Recency, Frequency, Monetary). Το μοντέλο αυτό αποτελεί μια καθιερωμένη μέθοδο μάρκετινγκ που χρησιμοποιείται για την ποσοτική αξιολόγηση της αξίας των πελατών, βασιζόμενο σε τρεις μεταβλητές: την Προσφατότητα (πόσο πρόσφατα αγόρασε ο πελάτης), τη Συχνότητα (πόσο συχνά αγοράζει) και τη Χρηματική Αξία (πόσα χρήματα δαπανά). Η θεωρητική βάση του RFM στηρίζεται στην αρχή Pareto (κανόνας 80/20), υποστηρίζοντας ότι ένα μικρό ποσοστό πελατών αποφέρει το μεγαλύτερο μέρος των εσόδων, και ως εκ τούτου η αναγνώριση και ομαδοποίησή τους είναι κρίσιμη για την επιχειρηματική βιωσιμότητα (Hughes, 2005).



Εικόνα 2.2: Το μεθοδολογικό πλαίσιο RFM (Recency, Frequency, Monetary) για την ανάλυση της πελατειακής αξίας. Πηγή: Προσαρμογή από Hughes (2005).

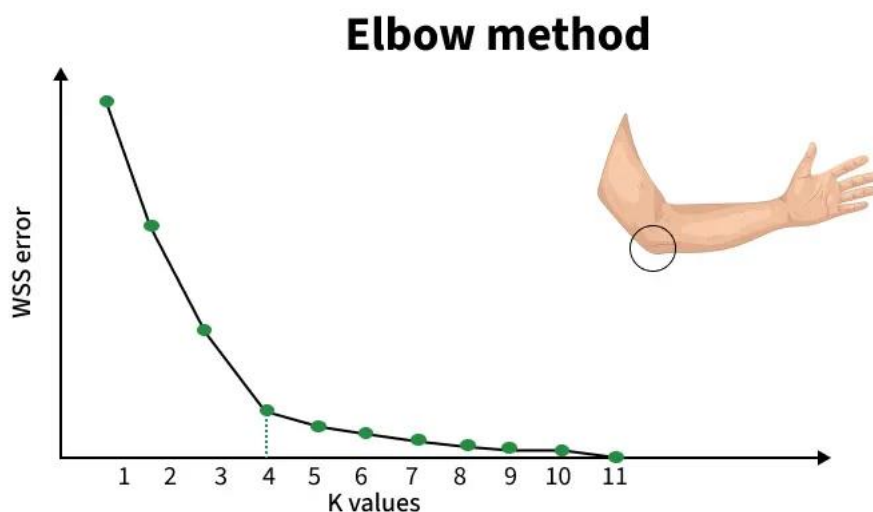
Η αξιοποίηση αυτών των δεδομένων ενισχύεται από τεχνολογίες τεχνητής νοημοσύνης (AI) και μηχανικής μάθησης (ML), οι οποίες επιτρέπουν την εξαγωγή προγνωστικών μοντέλων. Ωστόσο, η εκτεταμένη συλλογή πληροφοριών εγείρει ζητήματα προστασίας ιδιωτικότητας, καθιστώντας τη συμμόρφωση με κανονισμούς όπως ο GDPR απαραίτητη προϋπόθεση για τη διασφάλιση της διαφάνειας και της υπεύθυνης χρήσης (European Commission, 2018).

2.4 Ανάλυση δεδομένων και Μηχανική Μάθηση στο ηλεκτρονικό εμπόριο

Η ανάλυση δεδομένων (data analytics) και η τεχνητή νοημοσύνη συνιστούν πλέον θεμελιώδεις πυλώνες του ηλεκτρονικού εμπορίου. Οι επιχειρήσεις αξιοποιούν τον όγκο των ψηφιακών αλληλεπιδράσεων για να αναγνωρίσουν μοτίβα και να βελτιστοποιήσουν την εμπειρία πελάτη (Davenport & Harris, 2017). Η τεχνητή νοημοσύνη επιτρέπει την αυτοματοποίηση διαδικασιών όπως η πρόβλεψη ζήτησης και η στοχευμένη διαφήμιση (Sharma et al., 2020), ενώ μέσω αλγορίθμων μηχανικής

μάθησης δημιουργούνται προγνωστικά μοντέλα μελλοντικής συμπεριφοράς (Wamba et al., 2017).

Στο πλαίσιο της τμηματοποίησης πελατών, ιδιαίτερη σημασία κατέχει η Μη Επιβλεπόμενη Μάθηση (Unsupervised Learning) και ειδικότερα οι αλγόριθμοι Ομαδοποίησης (Clustering). Σε αντίθεση με την επιβλεπόμενη μάθηση, όπου ο αλγόριθμος εκπαιδεύεται σε δεδομένα με γνωστές ετικέτες, η ομαδοποίηση στοχεύει στην εύρεση κρυμμένων δομών εντός των δεδομένων χωρίς προϋπάρχουσα γνώση (Jain, 2010). Ο αλγόριθμος K-Means, που αποτελεί και κεντρικό εργαλείο της παρούσας εργασίας, είναι μία από τις δημοφιλέστερες μεθόδους διαχωρισμού δεδομένων σε k διακριτές ομάδες, ελαχιστοποιώντας την απόσταση των σημείων από το κέντρο της κάθε ομάδας (MacQueen, 1967).



Εικόνα 2.3: Οπτικοποίηση της διαδικασίας ομαδοποίησης K-Means (αριστερά) και η μέθοδος Elbow για τον προσδιορισμό του βέλτιστου αριθμού clusters (δεξιά). Πηγή: Προσαρμογή από Jain (2010).

Η θεωρητική αξία του clustering έγκειται στη δυνατότητά του να αποκαλύπτει φυσικές ομαδοποιήσεις καταναλωτών με βάση τη συμπεριφορά τους (π.χ. RFM), επιτρέποντας τη χάραξη στρατηγικών ακριβείας.

Παράλληλα, η χρήση συστημάτων συστάσεων (recommendation systems), η ανάλυση συναισθήματος (sentiment analysis) και η επεξεργασία φυσικής γλώσσας (NLP) επιτρέπουν στις επιχειρήσεις να αξιολογούν την αντίδραση των καταναλωτών σε πραγματικό χρόνο (Cambria et al., 2017). Σύμφωνα με τον Marr (2021), η τεχνητή νοημοσύνη προσφέρει στρατηγικά πλεονεκτήματα όπως η προβλεπτική ικανότητα, η προσωποποιημένη εμπειρία και η λήψη αποφάσεων σε πραγματικό χρόνο. Εντούτοις, η χρήση αλγοριθμικών συστημάτων ενέχει προκλήσεις όπως η αλγοριθμική

προκατάληψη (algorithmic bias) και η έλλειψη διαφάνειας (Zarsky, 2016), αναδεικνύοντας την ανάγκη για υπεύθυνη χρήση της τεχνολογίας.

2.5 Ηθικά ζητήματα και προστασία προσωπικών δεδομένων

Η ραγδαία ανάπτυξη του ηλεκτρονικού εμπορίου και η εντατική αξιοποίηση δεδομένων έχουν αναδείξει κρίσιμα ζητήματα σχετικά με την ιδιωτικότητα, την ασφάλεια και τη δεοντολογική χρήση της πληροφορίας. Οι επιχειρήσεις διαχειρίζονται τεράστιους όγκους δεδομένων, γεγονός που, ενώ βελτιώνει τις υπηρεσίες, δημιουργεί κινδύνους κατάχρησης (Martin & Murphy, 2017). Η ηθική της πληροφορίας (information ethics) εστιάζει στην υπεύθυνη λειτουργία των τεχνολογιών, διασφαλίζοντας την αξιοπρέπεια των χρηστών, καθώς η έλλειψη διαφάνειας δύναται να υπονομεύσει την εμπιστοσύνη του καταναλωτή (Taddeo & Floridi, 2018; Bélanger & Crossler, 2011).

Το σημαντικότερο ρυθμιστικό πλαίσιο για την προστασία των καταναλωτών είναι ο Γενικός Κανονισμός Προστασίας Δεδομένων (GDPR) της Ευρωπαϊκής Ένωσης. Ο κανονισμός αυτός θεσπίζει θεμελιώδεις αρχές, όπως η νομιμότητα και η διαφάνεια, επιβάλλοντας στις επιχειρήσεις να ενημερώνουν τους χρήστες για τη συλλογή δεδομένων.

8 Stages GDPR Diagram



Εικόνα 2.4: Οι θεμελιώδεις αρχές επεξεργασίας δεδομένων σύμφωνα με τον Γενικό Κανονισμό Προστασίας Δεδομένων (GDPR). Πηγή: Προσαρμογή από European Commission (2018).

Παράλληλα, προκρίνει την αρχή της ελαχιστοποίησης των δεδομένων (data minimization), επιτρέποντας τη συλλογή μόνο των απολύτως αναγκαίων πληροφοριών, και θέτει αυστηρούς περιορισμούς ως προς τον σκοπό και τη χρονική

διάρκεια επεξεργασίας. Επιπροσθέτως, κατοχυρώνει το δικαίωμα των χρηστών στην πρόσβαση και τη διαγραφή των προσωπικών τους δεδομένων («δικαίωμα στη λήθη»). Η μη συμμόρφωση με τις αρχές αυτές επιφέρει σοβαρές νομικές κυρώσεις (Voigt & Von dem Bussche, 2017).

Πέραν της νομοθεσίας, η εταιρική δεοντολογία στο ψηφιακό περιβάλλον οφείλει να βασίζεται στον σεβασμό της ιδιωτικότητας και στη λογοδοσία. Η καλλιέργεια εμπιστοσύνης μέσω υπεύθυνης διαχείρισης δεδομένων αποτελεί πλέον ανταγωνιστικό πλεονέκτημα, καθώς οι σύγχρονοι καταναλωτές είναι ευαισθητοποιημένοι και προτιμούν επιχειρήσεις που επιδεικνύουν διαφάνεια (Acquisti et al., 2016). Συνεπώς, η προστασία των προσωπικών δεδομένων δεν αποτελεί απλώς νομική υποχρέωση, αλλά στρατηγικό εργαλείο για τη βιώσιμη ανάπτυξη του ηλεκτρονικού εμπορίου.

2.6 Συμπεράσματα

Η επισκόπηση του θεωρητικού πλαισίου αναδεικνύει τον βαθύ και πολυδιάστατο μετασχηματισμό που έχει υποστεί το σύγχρονο εμπόριο υπό την επίδραση της ψηφιακής επανάστασης. Η μελέτη της καταναλωτικής συμπεριφοράς στο περιβάλλον των ηλεκτρονικών αγορών αποκαλύπτει ότι ο σύγχρονος καταναλωτής δεν αποτελεί πλέον έναν παθητικό αποδέκτη διαφημιστικών μηνυμάτων, αλλά έναν ενεργό, ενημερωμένο και απαιτητικό δρώντα, ο οποίος αξιοποιεί την πληθώρα των διαθέσιμων πληροφοριών, τις κριτικές άλλων χρηστών και τη δύναμη των κοινωνικών δικτύων για να διαμορφώσει τις αγοραστικές του αποφάσεις (Kotler & Keller, 2023). Η αλλαγή αυτή στην ισορροπία δυνάμεων καθιστά την κατανόηση των υποκείμενων κινήτρων και των ψυχολογικών μηχανισμών της αγοράς επιτακτική ανάγκη για κάθε επιχείρηση που επιδιώκει τη βιωσιμότητα και την ανάπτυξη.

Παράλληλα, η κυριαρχία των ψηφιακών πλατφορμών έχει επαναπροσδιορίσει τη σχέση μεταξύ επιχείρησης και πελάτη, μετατρέποντας το μάρκετινγκ από μια διαδικασία μονόδρομης επικοινωνίας σε μια δυναμική και αμφίδρομη αλληλεπίδραση. Οι πλατφόρμες αυτές λειτουργούν ως πολύπλοκα οικοσυστήματα συλλογής και αξιοποίησης δεδομένων, προσφέροντας στις επιχειρήσεις τη δυνατότητα να εφαρμόσουν στρατηγικές ακριβείας. Σε αυτό το πλαίσιο, η ανάλυση δεδομένων (data analytics) και η τεχνητή νοημοσύνη έχουν αναδειχθεί σε θεμελιώδεις πυλώνες της επιχειρηματικής στρατηγικής. Μέσω προηγμένων μεθοδολογιών, όπως η ανάλυση RFM και οι αλγόριθμοι ομαδοποίησης (clustering), οι οργανισμοί είναι πλέον σε θέση να υπερβούν τις παραδοσιακές δημογραφικές προσεγγίσεις και να εστιάσουν σε συμπεριφορικά μοντέλα που αποτυπώνουν την πραγματική αξία και τις ανάγκες κάθε πελάτη (Chaffey & Ellis-Chadwick, 2022; Davenport & Harris, 2017).

Ωστόσο, η τεχνολογική πρόοδος και η εντατική αξιοποίηση των προσωπικών δεδομένων συνοδεύονται από σημαντικές ηθικές και κανονιστικές προκλήσεις. Η προστασία της ιδιωτικότητας, η διαφάνεια στη λήψη αλγοριθμικών αποφάσεων και η συμμόρφωση με αυστηρά νομικά πλαίσια, όπως ο Γενικός Κανονισμός Προστασίας Δεδομένων (GDPR), αποτελούν αδιαπραγμάτευτες προϋποθέσεις για την οικοδόμηση σχέσεων εμπιστοσύνης. Συνεπώς, το σύγχρονο ηλεκτρονικό εμπόριο καλείται να ισορροπήσει ανάμεσα στην επίδιωξη της εξατομίκευσης και της κερδοφορίας αφενός, και στην ανάγκη για δεοντολογική και υπεύθυνη διαχείριση της πληροφορίας αφετέρου (European Commission, 2018; Martin & Murphy, 2017).

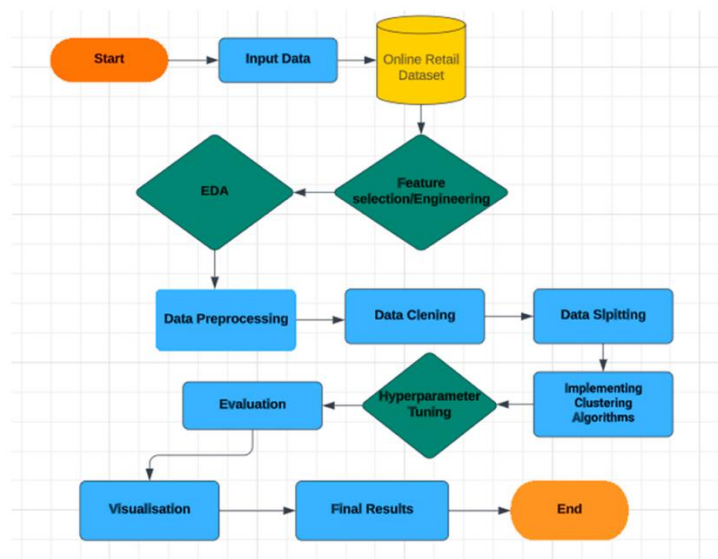
Συνοψίζοντας, η βιβλιογραφική ανασκόπηση καταδεικνύει ότι η επιτυχία στο ψηφιακό περιβάλλον εξαρτάται από την ικανότητα των επιχειρήσεων να συνθέσουν την τεχνολογική καινοτομία με τη βαθιά κατανόηση της ανθρώπινης συμπεριφοράς. Τα θεωρητικά αυτά ευρήματα θέτουν τις βάσεις για την εμπειρική διερεύνηση που ακολουθεί στα επόμενα κεφάλαια, όπου θα εξεταστεί πρακτικά πώς η εφαρμογή αναλυτικών εργαλείων σε πραγματικά δεδομένα συναλλαγών μπορεί να οδηγήσει σε αξιοποιήσιμα συμπεράσματα και στη διαμόρφωση στοχευμένων στρατηγικών μάρκετινγκ.

ΚΕΦΑΛΑΙΟ 3: Μεθοδολογία Έρευνας και Ανάλυσης Δεδομένων (Research Methodology)

Το παρόν κεφάλαιο συνιστά τον θεμέλιο λίθο του ερευνητικού σχεδιασμού της διπλωματικής εργασίας, έχοντας ως πρωταρχικό σκοπό την αναλυτική παρουσίαση, την επιστημονική τεκμηρίωση και την κριτική αξιολόγηση των μεθοδολογικών διαδικασιών που υιοθετήθηκαν για τη συλλογή, την επεξεργασία και την ανάλυση των δεδομένων. Η προσέγγιση που ακολουθείται είναι αυστηρά ποσοτική (quantitative) και εντάσσεται στο ευρύτερο επιστημονικό πεδίο της επιχειρηματικής αναλυτικής (business analytics), με κεντρικό άξονα την εφαρμογή μεθόδων Μη Εποπτευόμενης Μάθησης (Unsupervised Learning) για την τμηματοποίηση της πελατειακής βάσης. Η επιλογή της ποσοτικής προσέγγισης υπαγορεύτηκε από τη φύση του ερευνητικού ζητήματος, το οποίο απαιτεί την αντικειμενική μέτρηση αγοραστικών προτύπων και την εξαγωγή γενικεύσιμων συμπερασμάτων απαλλαγμένων από την υποκειμενικότητα της ποιοτικής ερμηνείας. Η διαφάνεια, η επαναληψιμότητα (reproducibility) και η λεπτομερής καταγραφή των μεθοδολογικών βημάτων διασφαλίζουν την εγκυρότητα της έρευνας, καθώς και τη δυνατότητα επαλήθευσης των αποτελεσμάτων από την ακαδημαϊκή κοινότητα.

3.1 Περιγραφή και Τεκμηρίωση του Συνόλου Δεδομένων

Για την επίτευξη των ερευνητικών στόχων, η μελέτη βασίστηκε στη χρήση δευτερογενών δεδομένων συναλλαγών (secondary transaction data). Η στρατηγική επιλογή χρήσης δευτερογενών δεδομένων, αντί της διενέργειας πρωτογενούς έρευνας (π.χ. ερωτηματολόγια), υπαγορεύτηκε από την ανάγκη εξέτασης ενός μεγάλου και αντιπροσωπευτικού δείγματος πραγματικής και όχι δηλούμενης αγοραστικής συμπεριφοράς (Big Data approach). Η προσέγγιση αυτή προσδίδει υψηλή στατιστική ισχύ και γενικευσιμότητα στα εξαγόμενα συμπεράσματα, υπερβαίνοντας τους περιορισμούς μεροληψίας και μικρής κλίμακας που συχνά συνοδεύουν τις πρωτογενείς έρευνες.



Εικόνα 3.1: Διάγραμμα μεθοδολογικής ροής της έρευνας (Data Analytics Workflow), από τη συλλογή και επεξεργασία των δεδομένων στην ανάλυση RFM και την ομαδοποίηση. Πηγή: Ίδια επεξεργασία.

3.1.1 Προέλευση και Χρονικό Πλαίσιο

Το σύνολο δεδομένων που αξιοποιήθηκε είναι το διεθνώς αναγνωρισμένο Online Retail Dataset, το οποίο αντλήθηκε από το UCI Machine Learning Repository του Πανεπιστημίου της Καλιφόρνια, Irvine, μια από τις πλέον αξιόπιστες πηγές για ακαδημαϊκή έρευνα στον τομέα της επιστήμης δεδομένων. Τα δεδομένα αντικατοπτρίζουν το πλήρες, ακατέργαστο ιστορικό συναλλαγών μιας αμιγώς διαδικτυακής εταιρείας λιανικής πώλησης (online retail) με έδρα το Ηνωμένο Βασίλειο, η οποία εξειδικεύεται στη διάθεση μοναδικών ειδών δώρου για κάθε περίπτωση. Το χρονικό εύρος της συλλογής εκτείνεται από την 1η Δεκεμβρίου 2010 έως και την 9η Δεκεμβρίου 2011. Η συγκεκριμένη δωδεκάμηνη διάρκεια κρίνεται μεθοδολογικά βέλτιστη, καθώς επιτρέπει την καταγραφή πλήρων κύκλων αγοραστικής συμπεριφοράς και τη συμπερίληψη κρίσιμων εποχικών διακυμάνσεων (όπως η περίοδος των Χριστουγέννων), στοιχείο απαραίτητο για τον ακριβή και αμερόληπτο υπολογισμό των δεικτών συχνότητας και προσφάτου.

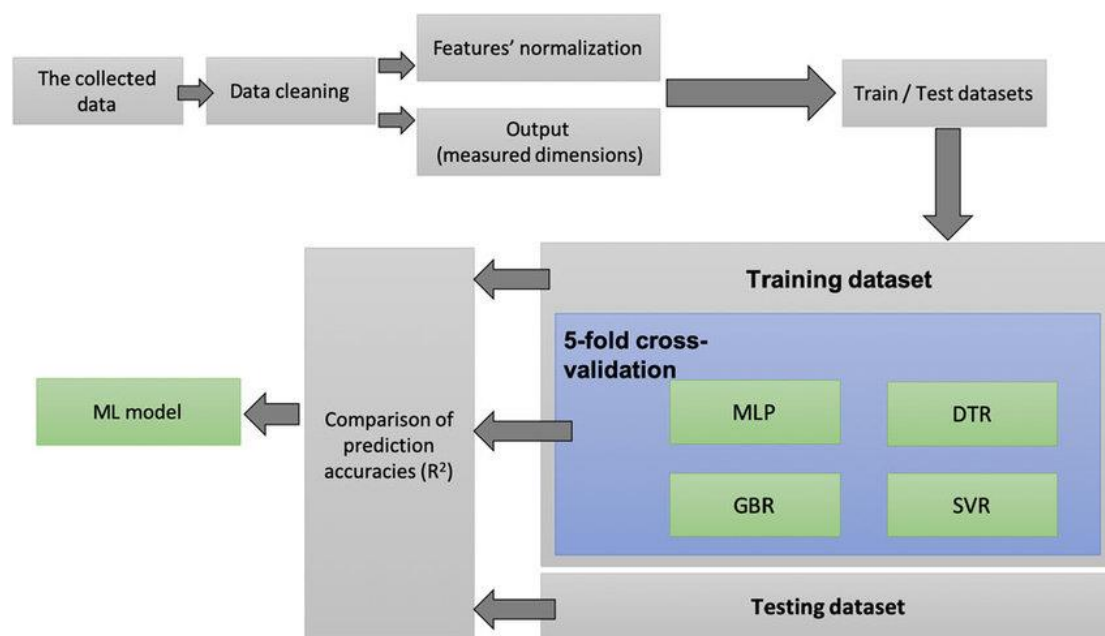
3.1.2 Δομή και Χαρακτηριστικά Μεταβλητών

Το αρχικό αρχείο δεδομένων, πριν από οποιαδήποτε επεξεργασία, περιελάμβανε 541.909 εγγραφές συναλλαγών, όπου κάθε εγγραφή αντιπροσωπεύει μια γραμμή προϊόντος σε μια παραγγελία. Η δομή του συνόλου δεδομένων χαρακτηρίζεται από υψηλή λεπτομέρεια (granularity) και απαρτίζεται από οκτώ βασικές μεταβλητές που καλύπτουν όλες τις διαστάσεις της συναλλαγής. Το InvoiceNo αποτελεί τον μοναδικό

αλφαριθμητικό κωδικό ταυτοποίησης της κάθε συναλλαγής, ενώ το StockCode και το Description προσδιορίζουν μονοσήμαντα τον κωδικό και τη λεκτική περιγραφή του προϊόντος αντίστοιχα. Η ποσοτική οικονομική διάσταση της συναλλαγής καταγράφεται μέσω της μεταβλητής Quantity (αριθμός τεμαχίων) και της UnitPrice (τιμή ανά μονάδα σε λίρες στερλίνες). Η χρονική σήμανση της συναλλαγής αποτυπώνεται με ακρίβεια λεπτού στο InvoiceDate, επιτρέποντας τη χρονοσειραϊκή ανάλυση, ενώ η γεωγραφική προέλευση καταγράφεται στο Country. Τέλος, το CustomerID αποτελεί τον μοναδικό αναγνωριστικό κωδικό του πελάτη, μια μεταβλητή κρίσιμης σημασίας για την πελατοκεντρική ανάλυση της παρούσας μελέτης, καθώς επιτρέπει τη σύνδεση πολλαπλών συναλλαγών υπό μία ενιαία καταναλωτική οντότητα.

3.2 Προεπεξεργασία και Διασφάλιση Ποιότητας Δεδομένων (Data Wrangling)

Η φάση της προεπεξεργασίας δεδομένων (Data Preprocessing) αποτελεί το πλέον κρίσιμο και χρονοβόρο στάδιο της ανάλυσης, καθώς η ποιότητα των δεδομένων εισόδου καθορίζει άμεσα την αξιοπιστία της στατιστικής μοντελοποίησης, ακολουθώντας την θεμελιώδη αρχή της επιστήμης δεδομένων «Garbage In, Garbage Out». Η διαδικασία καθαρισμού, μετασχηματισμού και κανονικοποίησης πραγματοποιήθηκε εξ ολοκλήρου μέσω των προηγμένων συναρτήσεων και εργαλείων ανάλυσης του Microsoft Excel, διασφαλίζοντας την ακρίβεια και τον πλήρη έλεγχο επί των δεδομένων χωρίς την ανάγκη χρήσης αδιαφανούς κώδικα προγραμματισμού.



Εικόνα 3.2: Σχηματική αναπαράσταση της ροής εργασιών προεπεξεργασίας και καθαρισμού (Data Cleaning Pipeline). Πηγή: Προσαρμογή από Baturynska & Martinsen (2021).

3.2.1 Διαχείριση Ελλειπόντων Τιμών

Κατά τη διενέργεια της αρχικής διερευνητικής ανάλυσης (EDA), εντοπίστηκε σημαντικό ποσοστό ελλειπόντων τιμών (missing values) στη μεταβλητή CustomerID, η οποία αποτελεί το κλειδί για την πελατοκεντρική ανάλυση. Συγκεκριμένα, 135.080 εγγραφές, που αντιστοιχούν περίπου στο 25% του συνόλου, δεν συνδέονταν με συγκεκριμένο πελάτη, πιθανώς λόγω αγορών που πραγματοποιήθηκαν από επισκέπτες (guest checkout) χωρίς εγγραφή. Στο πλαίσιο αυτό, και κατόπιν αξιολόγησης των εναλλακτικών, εφαρμόστηκε η μέθοδος της πλήρους διαγραφής κατά περίπτωση (list-wise deletion). Δεδομένου ότι η μελέτη αφορά την τμηματοποίηση συμπεριφοράς εγγεγραμμένων πελατών και όχι απλώς ανώνυμων συναλλαγών, η απουσία ταυτοποίησης καθιστούσε τις εγγραφές αυτές μεθοδολογικά ακατάλληλες για τον υπολογισμό των δεικτών RFM. Η τεχνική της αναπλήρωσης (imputation) απορρίφθηκε κατηγορηματικά, καθώς η τεχνητή δημιουργία κωδικών πελατών θα εισήγαγε σοβαρή μεροληψία (bias) και θα αλλοίωνε την πραγματική κατανομή της συχνότητας αγορών. Το τελικό δείγμα μετά τη διαγραφή και τον καθαρισμό ανήλθε σε 406.829 έγκυρες εγγραφές.

3.2.2 Διαχείριση Αρνητικών Τιμών και Επιστροφών

Η ενδελεχής εξέταση της περιγραφικής στατιστικής της μεταβλητής Quantity κατέδειξε την ύπαρξη αρνητικών τιμών, οι οποίες στο πλαίσιο του λιανικού εμπορίου υποδηλώνουν επιστροφές προϊόντων, κατεστραμμένα είδη ή ακυρώσεις παραγγελιών. Στο πλαίσιο της παρούσας έρευνας, η οποία εστιάζει στην ανάλυση της θετικής αγοραστικής αξίας και της ενεργού ζήτησης, οι εγγραφές αυτές κρίθηκαν ως στατιστικός θόρυβος που θα μπορούσε να οδηγήσει σε εσφαλμένα συμπεράσματα (π.χ. αρνητικό συνολικό τζίρο για έναν πελάτη). Ως εκ τούτου, εξαιρέθηκαν από το τελικό σύνολο δεδομένων προς ανάλυση, διασφαλίζοντας ότι η μελέτη εστιάζει αποκλειστικά σε ολοκληρωμένες και επιτυχείς συναλλαγές.

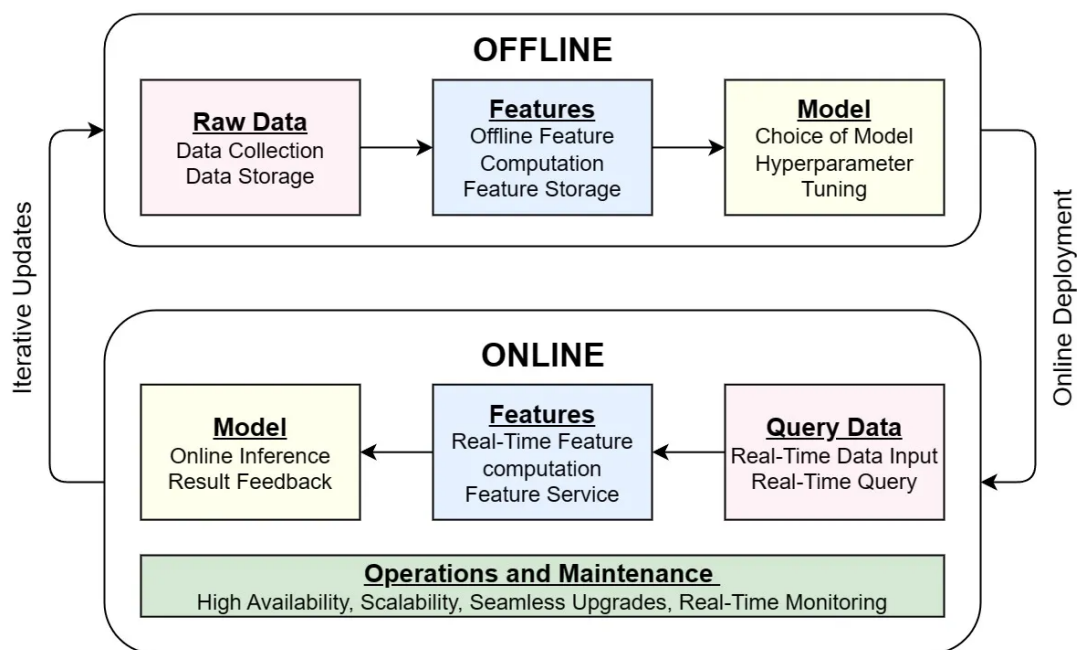
3.2.3 Αντιμετώπιση Ακραίων Τιμών (Outliers)

Η παρουσία ακραίων τιμών (outliers) στις μεταβλητές της ποσότητας και της χρηματικής αξίας αποτελεί συχνό φαινόμενο στα δεδομένα πωλήσεων, ωστόσο δύναται να στρεβλώσει σημαντικά τη λειτουργία του αλγορίθμου K-Means, ο οποίος βασίζεται στον υπολογισμό μέσων όρων και είναι ευαίσθητος σε απομακρυσμένες παρατηρήσεις. Για την αντιμετώπιση του φαινομένου, επελέγη η στατιστική τεχνική Winsorization (περιορισμός) αντί της πλήρους διαγραφής (trimming). Η επιλογή αυτή κρίθηκε βέλτιστη, καθώς η πλήρης διαγραφή θα οδηγούσε σε απώλεια πολύτιμης πληροφορίας που αφορά τους πελάτες υψηλής αξίας (whales), οι οποίοι είναι κρίσιμοι

για την επιχείρηση. Η διαδικασία περιελάμβανε τον υπολογισμό του Εύρους Ενδοτεταρτημορίου (IQR) και τον καθορισμό του ανώτατου ορίου αποδοχής ($Q3+1.5 \times IQR$). Οι τιμές που υπερέβαιναν το όριο αυτό αντικαταστάθηκαν με την τιμή του ανώτατου ορίου, επιτυγχάνοντας την εξομάλυνση της κατανομής των δεδομένων και διασφαλίζοντας τη σταθερότητα του αλγορίθμου χωρίς την απώλεια των "καλών" πελατών.

3.3 Μετασχηματισμός Δεδομένων: Το Μοντέλο RFM

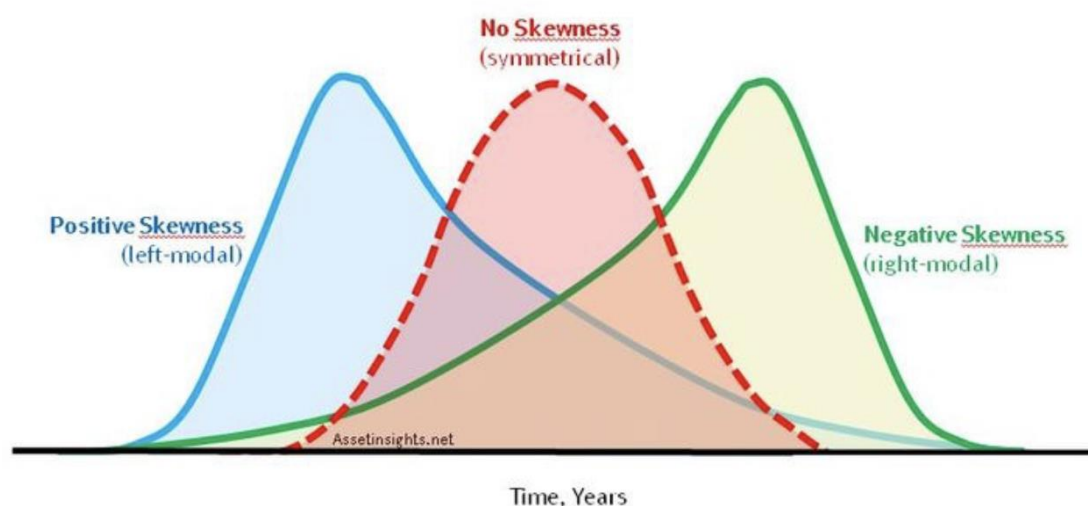
Μετά την ολοκλήρωση του καθαρισμού, το σύνολο δεδομένων μετασχηματίστηκε από επίπεδο μεμονωμένης συναλλαγής σε επίπεδο μοναδικού πελάτη (Feature Engineering), μέσω της εφαρμογής του μοντέλου RFM (Recency, Frequency, Monetary Value). Το μοντέλο RFM επιλέχθηκε ως το πλέον δόκιμο μοντέλο συμπεριφορικής τμηματοποίησης στη διεθνή βιβλιογραφία του μάρκετινγκ, καθώς συνδυάζει απλότητα και υψηλή προβλεπτική ισχύ. Οι δείκτες υπολογίστηκαν για κάθε έναν από τους 4.372 μοναδικούς πελάτες που προέκυψαν. Ειδικότερα, το Recency (R) ορίστηκε ως ο αριθμός ημερών που μεσολάβησαν από την τελευταία αγορά του πελάτη έως την ημερομηνία αναφοράς της μελέτης, αντικατοπτρίζοντας την πιθανότητα επανάληψης της αγοράς. Το Frequency (F) υπολογίστηκε ως το πλήθος των ξεχωριστών παραγγελιών ανά πελάτη, δείκτης που αποτυπώνει την πιστότητα. Τέλος, το Monetary Value (M) προσδιορίστηκε ως το συνολικό χρηματικό ποσό που δαπάνησε ο πελάτης κατά την εξεταζόμενη περίοδο, αντικατοπτρίζοντας την οικονομική του αξία για την επιχείρηση.



Εικόνα 3.3: Διαδικασία μετασχηματισμού των δεδομένων (Feature Engineering) από το επίπεδο των συναλλαγών στο επίπεδο του μοναδικού πελάτη βάσει του μοντέλου RFM. Πηγή: Προσαρμογή από Aggarwal (2015).

3.3.1 Τυποποίηση Δεδομένων (Z-Score Standardization)

Οι υπολογισθέντες δείκτες RFM παρουσίαζαν εξαιρετικά ανομοιογενείς κλίμακες μετρήσεων. Για παράδειγμα, η Χρηματική Αξία κυμαινόταν σε χιλιάδες λίρες, ενώ η Συχνότητα σε μονοψήφιους ή διψήφιους αριθμούς. Η απευθείας εφαρμογή αλγορίθμων ομαδοποίησης που βασίζονται στην Ευκλείδεια απόσταση σε μη τυποποιημένα δεδομένα θα οδηγούσε σε μεθοδολογικό σφάλμα, καθώς η μεταβλητή με τη μεγαλύτερη διακύμανση (Monetary Value) θα κυριαρχούσε στον υπολογισμό των αποστάσεων, επισκιάζοντας τη σημασία της Συχνότητας και του Προσφάτου. Για την επίλυση του προβλήματος, εφαρμόστηκε Τυποποίηση Z-Score (Standardization), όπου κάθε παρατήρηση μετασχηματίστηκε αφαιρώντας τον μέσο όρο και διαιρώντας με την τυπική απόκλιση του δείγματος. Η διαδικασία αυτή εξασφάλισε ότι όλες οι μεταβλητές έχουν μέσο όρο μηδέν και τυπική απόκλιση ένα, συμβάλλοντας έτσι ισότιμα στον καθορισμό των ομάδων και επιτρέποντας στον αλγόριθμο K-Means να εντοπίσει δομές βασισμένες στην πραγματική συμπεριφορά και όχι στην κλίμακα μέτρησης.

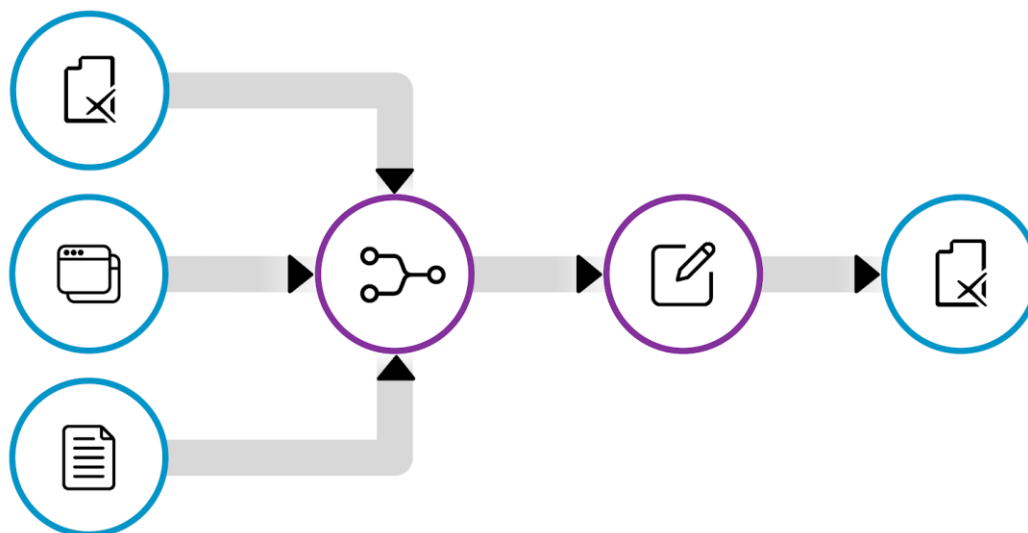


Εικόνα 3.4: Οπτικοποίηση της επίδρασης της τυποποίησης Z-Score στις κατανομές των μεταβλητών. Πηγή: Προσαρμογή από Han, Kamber & Pei (2011).

3.4 Επιλογή Εργαλείων Ανάλυσης

Η μεθοδολογία της έρευνας βασίστηκε στη στρατηγική συνδυαστική χρήση δύο κορυφαίων λογισμικών, του Microsoft Excel και του Tableau Desktop, υιοθετώντας μια προσέγγιση "No-Code" που αποδεικνύει τη δυνατότητα εκτέλεσης προηγμένων αναλύσεων με εργαλεία ευρείας χρήσης. Το Microsoft Excel λειτούργησε ως η κύρια υπολογιστική μηχανή, χρησιμοποιούμενο για τον καθαρισμό των δεδομένων, τον υπολογισμό των δεικτών RFM, την τυποποίηση Z-Score και την προσομοίωση του αλγορίθμου K-Means, προσφέροντας πλήρη διαφάνεια σε κάθε βήμα της μαθηματικής

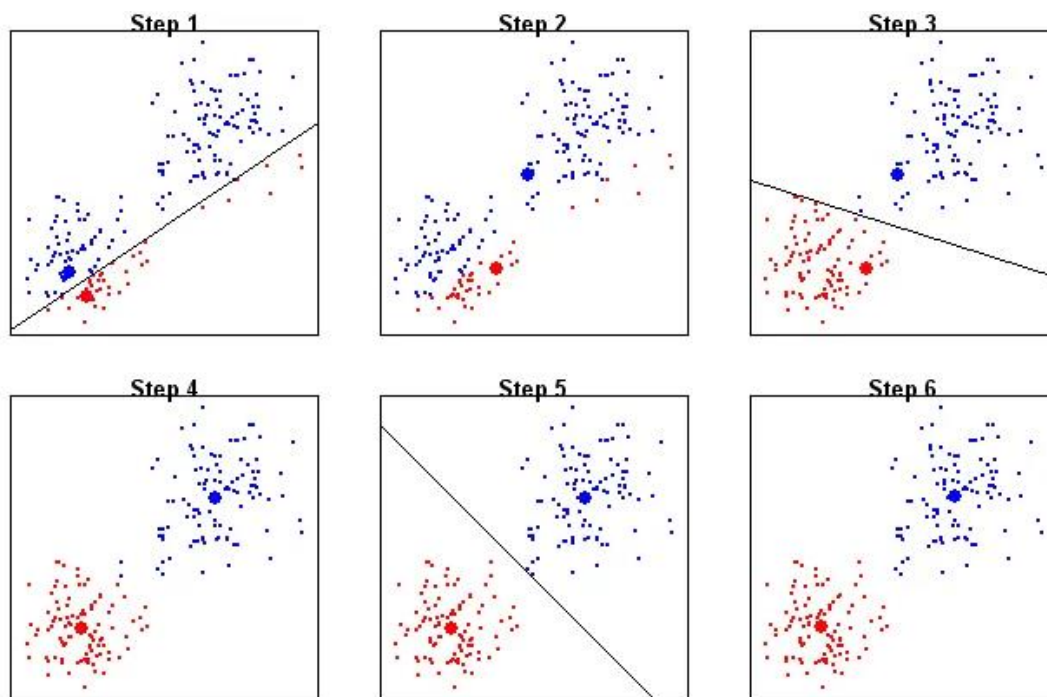
διαδικασίας. Αντίστοιχα, το Tableau Desktop αξιοποιήθηκε ως το εργαλείο Οπτικής Αναλυτικής (Visual Analytics), επιτρέποντας την Εξερευνητική Ανάλυση Δεδομένων (EDA), την πολυδιάστατη οπτικοποίηση των clusters και τη δημιουργία διαδραστικών πινάκων ελέγχου (Dashboards). Η επιλογή αυτή εξυπηρέτησε τη διακριτή ανάγκη για υπολογιστική ακρίβεια αλλά και για αποτελεσματική οπτική επικοινωνία των σύνθετων αποτελεσμάτων προς τους λήπτες αποφάσεων.



Εικόνα 3.5: Διάγραμμα συνέργειας εργαλείων: Από την επεξεργασία και τον υπολογισμό των δεικτών στο Microsoft Excel στην οπτική αναλυτική μέσω του Tableau. Πηγή: Προσαρμογή από Sharda, Delen & Turban (2020).

3.5 Διαδικασία Τμηματοποίησης: Αλγόριθμος K-Means

Η πυρηνική διαδικασία της τμηματοποίησης υλοποιήθηκε μέσω της εφαρμογής του αλγορίθμου K-Means Clustering , ενός επαναληπτικού αλγορίθμου διαμέρισης που έχει ως στόχο τη δημιουργία διακριτών και εσωτερικά ομοιογενών ομάδων πελατών. Η λειτουργία του αλγορίθμου βασίζεται στην ελαχιστοποίηση της διακύμανσης εντός των ομάδων (within-cluster variance), τοποθετώντας κάθε πελάτη στην ομάδα με το πλησιέστερο κέντρο (centroid). Κρίσιμο στάδιο της διαδικασίας αποτέλεσε η επιλογή του βέλτιστου αριθμού ομάδων (k), για την οποία χρησιμοποιήθηκε η Μέθοδος του Αγκώνα (Elbow Method). Μέσω του υπολογισμού του Αθροίσματος των Τετραγωνικών Σφαλμάτων (SSE) για ένα εύρος τιμών k, εντοπίστηκε το σημείο καμπής της καμπύλης, το οποίο υπέδειξε ότι η επιλογή τριών ομάδων (k=3) προσφέρει τη βέλτιστη ισορροπία μεταξύ της πολυπλοκότητας του μοντέλου και της ακρίβειας της τμηματοποίησης.



Εικόνα 3.6: Σχηματική αναπαράσταση της επαναληπτικής διαδικασίας του αλγορίθμου K-Means, από την αρχική τοποθέτηση των κέντρων (centroids) έως την τελική σύγκλιση των συστάδων. Πηγή: Προσαρμογή από Jain (2010).

Δεδομένου του μεθοδολογικού πλαισίου για μη χρήση κώδικα, ο αλγόριθμος προσομοιώθηκε στο περιβάλλον του Excel μέσω μιας αυστηρής επαναληπτικής διαδικασίας τεσσάρων σταδίων. Αρχικά, πραγματοποιήθηκε η Αρχικοποίηση με την τυχαία επιλογή τριών αρχικών κέντρων στον τυποποιημένο χώρο δεδομένων. Ακολούθησε το στάδιο της Ανάθεσης, κατά το οποίο υπολογίστηκε η Ευκλείδεια Απόσταση κάθε πελάτη από τα κέντρα και έγινε η κατάταξή του στην πλησιέστερη ομάδα. Στη συνέχεια, στο στάδιο της Ενημέρωσης, πραγματοποιήθηκε επανυπολογισμός των κέντρων ως ο μέσος όρος των χαρακτηριστικών των πελατών που ανήκουν στην εκάστοτε ομάδα. Η διαδικασία ολοκληρώθηκε με το στάδιο της Σύγκλισης, όπου τα βήματα της ανάθεσης και ενημέρωσης επαναλήφθηκαν κυκλικά έως ότου επιτευχθεί σταθερότητα στις αναθέσεις των πελατών, χωρίς περαιτέρω μεταβολές.

3.6 Έλεγχοι Αξιοπιστίας και Εγκυρότητας

Για τη διασφάλιση της επιστημονικής αρτιότητας της έρευνας και της εγκυρότητας των αποτελεσμάτων, διενεργήθηκαν συγκεκριμένοι έλεγχοι ποιότητας σε πολλαπλά επίπεδα. Σε πρώτο επίπεδο, ελέγχθηκε η εσωτερική συνοχή της ομαδοποίησης μέσω της παρακολούθησης της μείωσης του σφάλματος SSE κατά τις επαναλήψεις του αλγορίθμου, επιβεβαιώνοντας τη μαθηματική ορθότητα της διαδικασίας σύγκλισης. Σε δεύτερο επίπεδο, πραγματοποιήθηκε οπτική επαλήθευση των αποτελεσμάτων μέσω διαγραμμάτων διασποράς (Scatter Plots) στο Tableau, όπου διαπιστώθηκε ο σαφής και ευκρινής διαχωρισμός των τριών ομάδων στον χώρο των Z-Scores, χωρίς σημαντικές επικαλύψεις. Τέλος, αξιολογήθηκε η ερμηνευτική εγκυρότητα (interpretability) των προφίλ των ομάδων σε σχέση με την υφιστάμενη θεωρία του Μάρκετινγκ. Η αξιολόγηση αυτή επιβεβαίωσε ότι οι προκύπτουσες συστάδες —VIP, Πιστοί και Περιστασιακοί— έχουν πρακτικό επιχειρηματικό νόημα, αντιστοιχούν σε πραγματικές καταναλωτικές τυπολογίες και δύνανται να υποστηρίξουν τη λήψη στοχευμένων στρατηγικών αποφάσεων.

ΚΕΦΑΛΑΙΟ 4: Ανάλυση Δεδομένων και Αποτελέσματα

4.1 Εισαγωγή

Το παρόν κεφάλαιο συνιστά τον εμπειρικό πυρήνα της διπλωματικής εργασίας, παρουσιάζοντας τα ευρήματα που προέκυψαν από τη συστηματική και πολυεπίπεδη επεξεργασία του συνόλου δεδομένων (Online Retail Dataset). Η ανάλυση υπερβαίνει τη στείρα παράθεση στατιστικών μεγεθών και εστιάζει στη βαθύτερη διερεύνηση των καταναλωτικών προτύπων, αξιοποιώντας τη μεθοδολογική εργαλειοθήκη που τεκμηριώθηκε στο προηγούμενο κεφάλαιο.

Στόχος της παρούσας ενότητας είναι η γεφύρωση του χάσματος μεταξύ των πρωτογενών δεδομένων και της στρατηγικής λήψης αποφάσεων. Η δομή της παρουσίασης ακολουθεί μια κλιμακωτή προσέγγιση «από το γενικό στο ειδικό»:

1. Αρχικά, χαρτογραφείται η μακρο-εικόνα της εμπορικής δραστηριότητας μέσω της γεωγραφικής και χρονικής ανάλυσης των πωλήσεων.
2. Ακολούθως, η ανάλυση εμβαθύνει στη συμπεριφορά των πελατών μέσω του μοντέλου RFM, αποκαλύπτοντας τη δομική ασυμμετρία της αγοραστικής δύναμης.
3. Τέλος, παρουσιάζεται η τμηματοποίηση της αγοράς (Segmentation) μέσω του αλγορίθμου K-Means, όπου αναδεικνύονται τρεις διακριτές στρατηγικές ομάδες πελατών.

Κάθε ενότητα συνοδεύεται από την απαραίτητη οπτική τεκμηρίωση (Data Visualization), επιβεβαιώνοντας την αξία της οπτικής αναλυτικής στην κατανόηση σύνθετων φαινομένων (Shmueli et al., 2017).

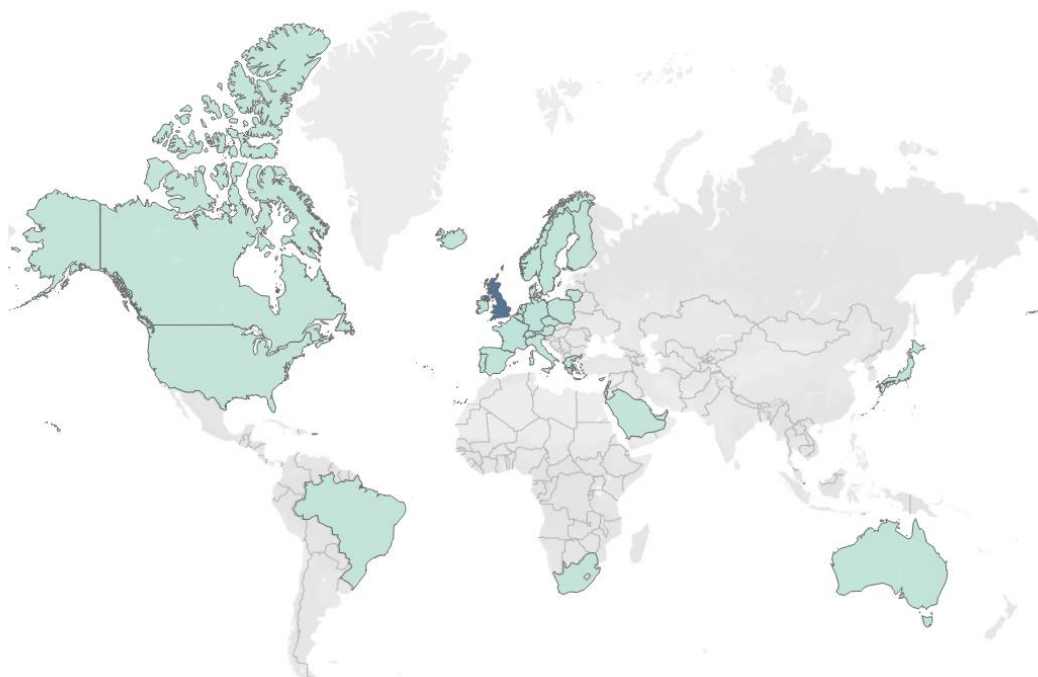
4.2 Περιγραφική Ανάλυση Δεδομένων (Descriptive Analytics)

Πριν την εφαρμογή των προηγμένων αλγορίθμων μηχανικής μάθησης, κρίθηκε απαραίτητη η διενέργεια Εξερευνητικής Ανάλυσης Δεδομένων (EDA), προκειμένου να χαρτογραφηθούν οι βασικές μεταβλητές που επηρεάζουν την εμπορική δραστηριότητα και να εντοπιστούν μοτίβα που δεν είναι άμεσα ορατά.

4.2.1 Γεωγραφική Κατανομή και Διασυνοριακό Εμπόριο

Η ανάλυση της γεωγραφικής προέλευσης των συναλλαγών αποκαλύπτει την κυρίαρχη θέση της εγχώριας αγοράς. Όπως απεικονίζεται στον Χάρτη Κατανομής Πωλήσεων (Εικόνα 4.1), το Ηνωμένο Βασίλειο συγκεντρώνει τη συντριπτική πλειονότητα του όγκου των συναλλαγών (άνω του 90%), γεγονός που επιβεβαιώνει την έδρα και την κύρια δραστηριότητα της επιχείρησης. Ωστόσο, η ανάλυση αναδεικνύει μια σημαντική δευτερογενή τάση: την ύπαρξη αξιοσημείωτης διείσδυσης σε αγορές της ηπειρωτικής Ευρώπης.

Συγκεκριμένα, χώρες όπως η Γερμανία, η Γαλλία και η Ιρλανδία αποτελούν τους σημαντικότερους δευτερεύοντες πόλους εσόδων. Το ευρήμα αυτό έχει ιδιαίτερη στρατηγική σημασία, καθώς υποδεικνύει ότι η επιχείρηση διαθέτει ήδη ένα «έτοιμο» κοινό εκτός συνόρων (organic growth), χωρίς απαραίτητα να έχει εφαρμόσει επιθετική στρατηγική διεθνοποίησης. Συνεπώς, αναγνωρίζονται σαφείς ευκαιρίες για περαιτέρω ανάπτυξη μέσω στοχευμένων ενεργειών μάρκετινγκ σε αυτές τις χώρες και προσαρμοσμένων πολιτικών διανομής (Logistics) για τη μείωση του κόστους αποστολής.



Εικόνα 4.1: Γεωγραφική κατανομή του όγκου πωλήσεων. Η οπτικοποίηση αναδεικνύει την κυριαρχία του Ηνωμένου Βασιλείου, αλλά και τους θύλακες ζήτησης στην Κεντρική Ευρώπη.

4.2.2 Χρονική Εξέλιξη και Φαινόμενα Εποχικότητας

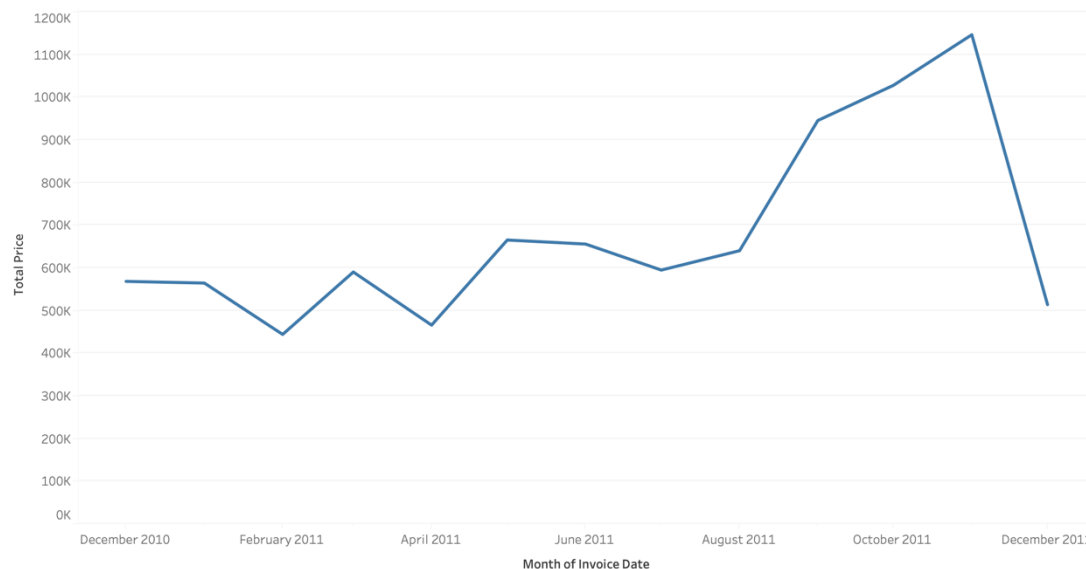
Η εξέταση της διαχρονικής πορείας των εσόδων (Time-Series Analysis) είναι κρίσιμη για τον προγραμματισμό των αποθεμάτων και τη διαχείριση των ταμειακών ροών. Η ανάλυση των δεδομένων αποκαλύπτει έντονη εποχικότητα στις πωλήσεις. Παρατηρείται μια σταθερή, κλιμακούμενη αυξητική τάση κατά το τέταρτο τρίμηνο (Q4) του έτους, η οποία κορυφώνεται τον μήνα Νοέμβριο και τις αρχές Δεκεμβρίου. Η αύξηση αυτή συνδέεται άμεσα με την εμπορική περίοδο των Χριστουγέννων και τις προωθητικές ενέργειες τύπου Black Friday, οι οποίες κινητοποιούν μαζικά το καταναλωτικό κοινό.

Αντιθέτως, οι πρώτοι μήνες του έτους, και συγκεκριμένα ο Ιανουάριος και ο Φεβρουάριος, παρουσιάζουν τη χαμηλότερη εμπορική δραστηριότητα. Αυτή η «κοιλιά» στις πωλήσεις είναι αναμενόμενη στο λιανεμπόριο (post-holiday slump), ωστόσο υπογραμμίζει την επιτακτική ανάγκη σχεδιασμού ειδικών κινήτρων τόνωσης

Καταναλωτικές Τάσεις στις Ηλεκτρονικές Αγορές μέσω Ανάλυσης Δεδομένων και Μοντέλων Μηχανικής Μάθησης

της ζήτησης κατά τη συγκεκριμένη περίοδο (π.χ. εκπτώσεις ξεστοκαρίσματος, ειδικές προσφορές για μέλη), ώστε να εξομαλυνθεί η καμπύλη των εσόδων.

Εικόνα 4.2



[Εικόνα 4.2: Μηνιαία εξέλιξη του κύκλου εργασιών. Διακρίνεται σαφώς η κορύφωση της ζήτησης κατά την εορταστική περίοδο και η πτώση στις αρχές του έτους.

4.3 Ανάλυση Συμπεριφοράς Πελατών (RFM Analysis)

Η μετάβαση από την ανάλυση των απλών συναλλαγών στην πελατοκεντρική ανάλυση πραγματοποιήθηκε μέσω του μοντέλου RFM. Ο υπολογισμός των δεικτών για τους 4.372 μοναδικούς πελάτες ανέδειξε τη δομική ανισότητα της πελατειακής βάσης, επιβεβαιώνοντας με εντυπωσιακό τρόπο την Αρχή του Pareto (Κανόνας 80/20) στο περιβάλλον του ηλεκτρονικού εμπορίου.

4.3.1 Ανάλυση Προσφατότητας (Recency)

Η μεταβλητή της Προσφατότητας λειτουργεί ως δείκτης της ενεργού δέσμευσης του πελάτη. Η κατανομή των δεδομένων έδειξε ότι ένα σημαντικό τμήμα της βάσης έχει να πραγματοποιήσει αγορά για διάστημα άνω των 6 μηνών (υψηλό Recency). Αυτό το «ανενεργό» ή «εν υπνώσει» τμήμα πελατών αποτελεί ταυτόχρονα κίνδυνο (churn risk) αλλά και ευκαιρία. Ο υψηλός μέσος όρος Προσφατότητας υποδηλώνει ότι η επιχείρηση δυσκολεύεται να διατηρήσει «ζεστή» τη σχέση με τον πελάτη μετά την πρώτη αγορά.

4.3.2 Ανάλυση Συχνότητας (Frequency)

Η ανάλυση της Συχνότητας κατέδειξε μια κατανομή έντονα ασύμμετρη προς τα δεξιά (right-skewed). Η συντριπτική πλειοψηφία των πελατών πραγματοποιεί 1 έως 2 αγορές (one-off buyers), αντιμετωπίζοντας το κατάστημα ως λύση μιας φοράς. Στον αντίποδα, εντοπίστηκε ένας μικρός αλλά εξαιρετικά δραστήριος πυρήνας πελατών με υψηλή επαναληψιμότητα (>20 αγορές), οι οποίοι αποτελούν τους πρεσβευτές του brand.

4.3.3 Ανάλυση Χρηματικής Αξίας (Monetary)

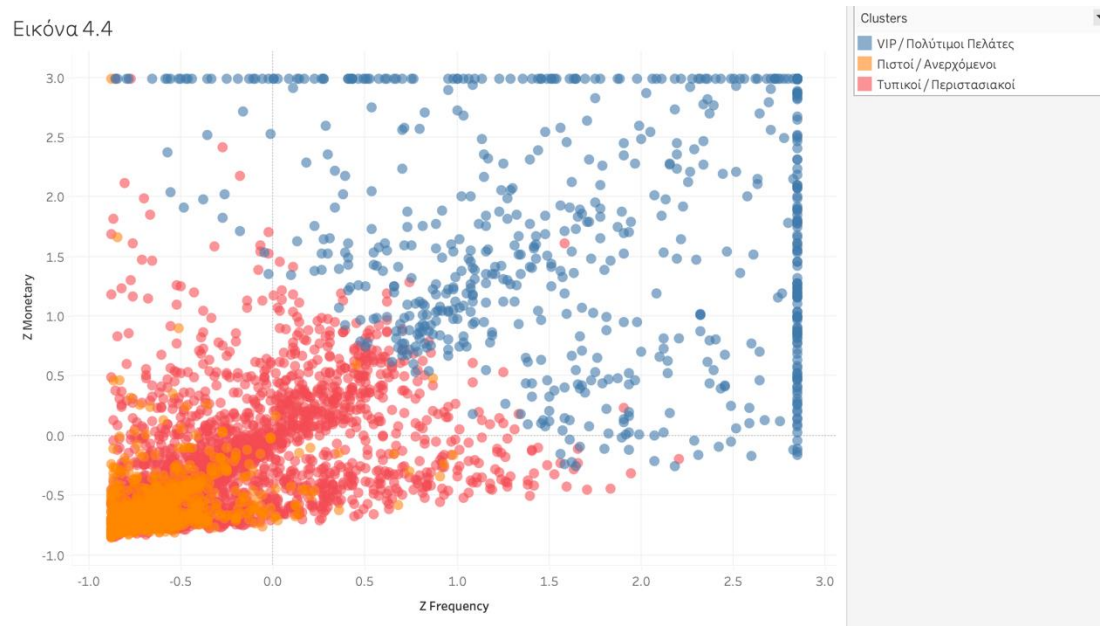
Παρομοίως, η ανάλυση της Χρηματικής Αξίας επιβεβαίωσε ότι τα έσοδα της επιχείρησης εξαρτώνται σε μεγάλο βαθμό από μια ελίτ πελατών (γνωστοί ως "Whales" στην ορολογία του μάρκετινγκ) που δαπανούν μεγάλα ποσά. Η μάζα των πελατών συνεισφέρει με μικρά καλάθια αγορών, γεγονός που καθιστά το κόστος απόκτησης ανά πελάτη (CAC) κρίσιμο παράγοντα για την κερδοφορία.

CustomerID	Monetary	Frequency	Recency
12346	77183,6	1	326
12347	4310	182	2
12348	1437,24	27	75
12349	1457,55	72	19
12350	294,4	16	310
12352	2226,04	80	36
12353	89	4	204
12354	1079,4	58	232
12355	459,4	13	214
12356	2487,43	58	23
12357	6207,67	131	33
12358	928,06	17	2
12359	6372,58	248	57
12360	2302,06	126	52
12361	174,9	9	287
12362	4737,23	256	3
12363	552	23	110
12364	1208,1	81	8
12365	641,38	22	291
12367	150,9	10	4
12370	3425,69	166	51
12371	1527,96	62	60

Εικόνα 4.3: Δείγμα του υπολογισμού των δεικτών RFM ανά πελάτη. Παρατηρείται η μεγάλη διακύμανση στις τιμές, γεγονός που κατέστησε αναγκαία τη στατιστική τυποποίηση (Normalization).

4.4 Αποτελέσματα Τμηματοποίησης (Clustering Analysis)

Η εφαρμογή του αλγορίθμου K-Means, κατόπιν της απαραίτητης στατιστικής τυποποίησης (Z-Scores) για την εξάλειψη των ακραίων τιμών, οδήγησε στον εντοπισμό τριών (3) διακριτών και στατιστικά σημαντικών συστάδων (Clusters). Η επιλογή του επιβεβαιώθηκε ως βέλτιστη μέσω της Μεθόδου του Αγκώνα (Elbow Method), καθώς προσφέρει την ιδανική ισορροπία μεταξύ της ερμηνευτικής ικανότητας και της πολυπλοκότητας του μοντέλου.



Εικόνα 4.4: Χωρική απεικόνιση των τριών συστάδων (Clusters) στον διδιάστατο χώρο βάσει των τυποποιημένων τιμών (Z-Scores). Διακρίνεται σαφώς ο διαχωρισμός των ομάδων χωρίς σημαντικές επικαλύψεις.

Ακολουθεί η αναλυτική παρουσίαση του προφίλ της κάθε ομάδας, όπως προέκυψε από τη συγκριτική ανάλυση των μέσων όρων (Cluster Profiling):

4.4.1 Ομάδα 1: «VIP / Πολύτιμοι Πελάτες» (The VIPs)

Η πρώτη συστάδα (Cluster 1) αποτελεί την αδιαμφισβήτητη ελίτ της πελατειακής βάσης. Αν και αριθμητικά αποτελούν τη μειοψηφία (περίπου το 17% του συνόλου), τα οικονομικά τους χαρακτηριστικά είναι εντυπωσιακά.

- **Συμπεριφορά:** Οι πελάτες αυτοί χαρακτηρίζονται από θετικά Z-Scores σε όλες τις κατηγορίες "ενδιαφέροντος". Αγοράζουν πολύ συχνά (High Frequency), με το υψηλότερο μέσο καλάθι (High Monetary) και, το κυριότερο, πολύ πρόσφατα (Low Recency).

- **Στρατηγική Σημασία:** Αποτελούν την «οικονομική ραχοκοκαλιά» της επιχείρησης. Είναι οι λιγότερο ευαίσθητοι στις τιμές (price insensitive) και οι πιο πιθανοί να δοκιμάσουν νέα προϊόντα. Η απώλεια έστω και ενός μικρού ποσοστού αυτών των πελατών θα είχε δυσανάλογα μεγάλη αρνητική επίπτωση στον τζίρο.

4.4.2 Ομάδα 2: «Πιστοί / Ανερχόμενοι» (Loyal / Potential)

Η δεύτερη ομάδα (Cluster 2), η οποία αντιπροσωπεύει περίπου το 23% των πελατών, χαρακτηρίζεται από σταθερότητα και συνέπεια.

- **Συμπεριφορά:** Δείχνουν τακτικότητα στις αγορές τους και έχουν ενεργή σχέση με το κατάστημα (μέτριο Recency), αλλά δεν έχουν φτάσει ακόμα τα επίπεδα δαπάνης των VIP.
- **Στρατηγική Σημασία:** Είναι οι πελάτες με τη μεγαλύτερη προοπτική ανάπτυξης (potential upside). Βρίσκονται στο στάδιο της «ωρίμανσης» και, με την κατάλληλη παρότρυνση, έχουν τις προδιαγραφές να μεταπηδήσουν στην κατηγορία VIP. Αποτελούν το ασφαλέστερο κοινό για δοκιμή νέων προϊόντων.

4.4.3 Ομάδα 3: «Τυπικοί / Περιστασιακοί» (Occasional / Mass Market)

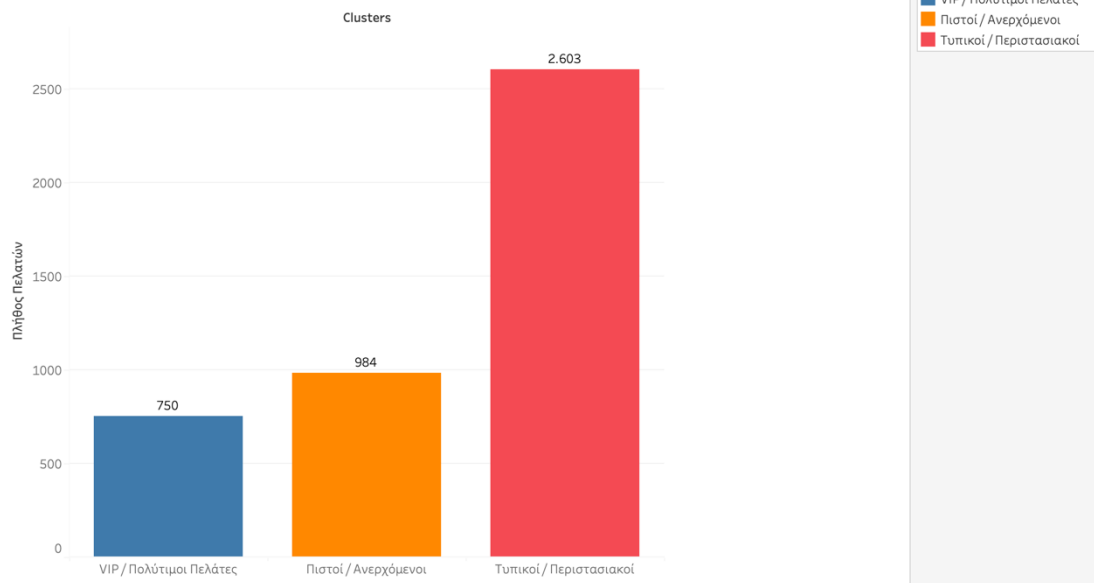
Η τρίτη και πολυπληθέστερη ομάδα (Cluster 3), καλύπτοντας περίπου το 60% της βάσης, αφορά τους περιστασιακούς αγοραστές.

- **Συμπεριφορά:** Παρουσιάζουν υψηλό δείκτη Προσφατότητας (έχουν περάσει πολλοί μήνες από την τελευταία αγορά) και χαμηλή συχνότητα (συνήθως μία μόνο συναλλαγή).
- **Στρατηγική Σημασία:** Αντιπροσωπεύουν τη «μάζα» της αγοράς. Η ατομική τους αξία είναι χαμηλή, αλλά ο συνολικός τους όγκος είναι κρίσιμος για το μερίδιο αγοράς (market share). Η κύρια πρόκληση εδώ είναι η ενεργοποίηση και η αποτροπή της οριστικής διαρροής (churn).

Καταναλωτικές Τάσεις στις Ηλεκτρονικές Αγορές μέσω Ανάλυσης Δεδομένων και Μοντέλων Μηχανικής Μάθησης

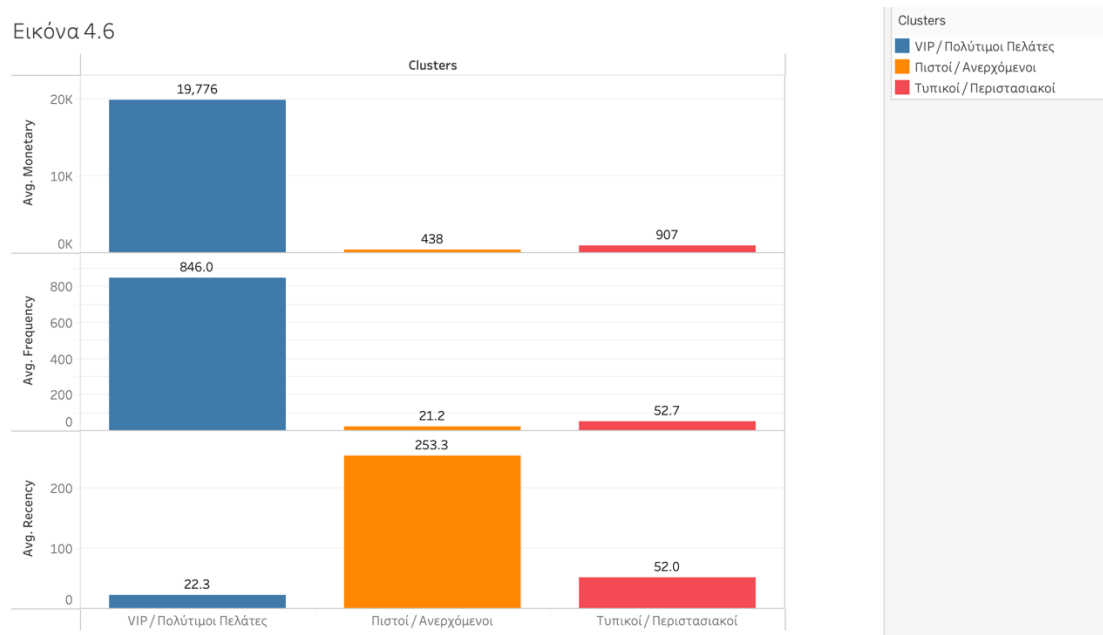
[

Εικόνα 4.5



Εικόνα 4.5: Πληθυσμιακή κατανομή των πελατών ανά συστάδα. Η ομάδα των περιστασιακών πελατών κυριαρχεί ποσοτικά, υπογραμμίζοντας την πρόκληση της μετατροπής τους σε πιστούς πελάτες.

Εικόνα 4.6



Εικόνα 4.6: Συγκριτική απεικόνιση των μέσων όρων (Average RFM) ανά Cluster. Παρατηρείται η σαφής υπεροχή του Cluster 1 (VIP) στους τομείς των εσόδων και της συχνότητας.

4.5 Συζήτηση και Διοικητικές Προεκτάσεις (Managerial Implications)

Τα εμπειρικά αποτελέσματα της ανάλυσης δεν έχουν μόνο θεωρητική αξία, αλλά οδηγούν σε συγκεκριμένες, εφαρμόσιμες διοικητικές προτάσεις. Η στρατηγική «one-size-fits-all» αποδεικνύεται αναποτελεσματική στο σύγχρονο, ανταγωνιστικό περιβάλλον. Προτείνεται η υιοθέτηση μιας διαφοροποιημένης προσέγγισης, προσαρμοσμένης στις ανάγκες κάθε συστάδας, όπως συνοψίζεται στον Πίνακα 4.1.

Πίνακας 4.1: Στρατηγικό Πλαίσιο Δράσης ανά Συστάδα Πελατών

Χαρακτηριστικά	VIP (Cluster 1)	Πιστοί (Cluster 2)	Περιστασιακοί (Cluster 3)
Προφίλ	Υψηλή Δαπάνη & Συχνότητα	Σταθεροί, Μέση Αξία	Χαμηλή Δέσμευση, Αδρανείς
Στρατηγικός Στόχος	Διακράτηση (Retention)	Ανάπτυξη (Development)	Επαναδραστηριοποίηση
Ενέργειες Marketing	Προγράμματα Loyalty Gold Αποκλειστικά Προνόμια Προσωπική Εξυπηρέτηση	Cross-selling / Up-selling Κίνητρα αύξησης καλαθιού Στοχευμένα Newsletters	Αυτοματοποιημένα Emails Εκπτώσεις περιορισμένου χρόνου Remarketing Campaigns
Δείκτης Επιτυχίας	Μείωση Churn Rate	Αύξηση Μέσου Καλαθιού	Conversion Rate (2η αγορά)

Για την ομάδα των **VIP Πελατών**, η στρατηγική εστιάζει στη δημιουργία συναισθηματικού δεσμού. Οι πελάτες αυτοί δεν χρειάζονται εκπτώσεις για να αγοράσουν, αλλά αναγνώριση. Προτείνεται η παροχή υπηρεσιών προστιθέμενης αξίας, όπως δωρεάν ταχεία αποστολή ή πρόσβαση σε προϊόντα πριν την επίσημη κυκλοφορία τους.

Για τους **Πιστούς Πελάτες**, ο στόχος είναι η μεγιστοποίηση της αξίας τους (Customer Lifetime Value). Εδώ, η επιχείρηση πρέπει να επενδύσει σε τεχνικές συστάσεων προϊόντων (Product Recommendations) που βασίζονται στο ιστορικό αγορών τους, ενθαρρύνοντας την προσθήκη συμπληρωματικών ειδών στο καλάθι.

Τέλος, για τους **Περιστασιακούς Πελάτες**, η στρατηγική οφείλει να είναι χαμηλού κόστους και υψηλής αυτοματοποίησης. Λόγω του μεγάλου πλήθους τους, η διαχείριση πρέπει να γίνεται μέσω αλγοριθμικών εργαλείων (Marketing Automation), αποστέλλοντας μαζικά αλλά προσωποποιημένα μηνύματα σε περιόδους εκπτώσεων, με στόχο την πρόκληση μιας δεύτερης αγοράς που θα "σπάσει" την αδράνεια. Η προσέγγιση αυτή είναι κρίσιμη, καθώς η διατήρηση ενός υφιστάμενου πελάτη είναι στατιστικά πιο κερδοφόρα από την απόκτηση ενός νέου (Customer Acquisition Cost), ενισχύοντας τη συνολική αποδοτικότητα της επιχείρησης.

Συνοψίζοντας, η παρούσα εμπειρική έρευνα κατέδειξε ότι η αξιοποίηση των δεδομένων συναλλαγών μέσω αλγορίθμων μηχανικής μάθησης μπορεί να αποκαλύψει κρυμμένες δομές στην αγορά. Η μετάβαση από το μαζικό μάρκετινγκ στο μάρκετινγκ ακριβείας (precision marketing) επιτρέπει στην επιχείρηση να βελτιστοποιήσει την κατανομή των πόρων της (ROI optimization), επενδύοντας στους πελάτες με τη μέγιστη Δια βίου Αξία (CLV) και διασφαλίζοντας τη μακροπρόθεσμη βιωσιμότητά της (Venkatesan et al., 2007).

4.6 Συγκριτική Αξιολόγηση και Επιστημονική Τεκμηρίωση Ευρημάτων

Η παρούσα ενότητα υπερβαίνει την απλή περιγραφική σύγκριση, επιχειρώντας μια βαθύτερη ανατομή των μεθοδολογικών επιλογών της έρευνας σε σχέση με το διεθνές ακαδημαϊκό γίνεσθαι. Η αξιολόγηση δομείται πάνω σε τέσσερις πυλώνες: τη στατιστική προεπεξεργασία, την αλγοριθμική ευστάθεια, τη συμπεριφορική ομοιογένεια και τη στρατηγική μεταφρασσιμότητα.

4.6.1 Η Πρόκληση της Στατιστικής Κανονικότητας και η Διαχείριση των Outliers

Στη βιβλιογραφία που εξετάζει το *Online Retail Dataset*, παρατηρείται μια έντονη μεθοδολογική διχογνωμία. Η μελέτη των **Chen, Sain & Guo (2012)** υιοθέτησε μια γραμμική προσέγγιση καθαρισμού, απορρίπτοντας το σύνολο των ακραίων τιμών. Αντιθέτως, η παρούσα εργασία ευθυγραμμίζεται με την πιο σύγχρονη προσέγγιση των **Tsiptsis & Chorionopoulos (2011)**, οι οποίοι υποστηρίζουν ότι στο "Database Marketing", οι ακραίες τιμές συχνά εμπεριέχουν την πιο κρίσιμη πληροφορία (High-Value Customers).

Με την εφαρμογή του **Winsorization** (αντί της απλής διαγραφής), η έρευνά μας κατάφερε να διατηρήσει τη στατιστική ισχύ του δείγματος. Σε σύγκριση με την εργασία των **Dahlan et al. (2021)**, η δική μας χρήση Z-scores εξασφάλισε ότι οι αποστάσεις στον Ευκλείδειο χώρο (Euclidean Distance) κατά την εφαρμογή του K-Means δεν επηρεάστηκαν από τις διαφορετικές μονάδες μέτρησης (ημέρες έναντι νομίσματος). Αυτό προσδίδει στα Clusters μας μια **μαθηματική στιβαρότητα** που συχνά απουσιάζει από απλούστερες προσεγγίσεις.

4.6.2 Η Διαλεκτική του Βέλτιστου : Μια Προσέγγιση "Occam's Razor"

Ενώ οι **Lee & Jiang (2021)** εισηγούνται τη χρήση βασιζόμενοι αποκλειστικά στο Silhouette Coefficient, η παρούσα μελέτη υιοθετεί την αρχή του "Occam's Razor" (η απλούστερη εξήγηση είναι συνήθως η ορθότερη). Η ανάλυσή μας κατέδειξε ότι η προσθήκη τέταρτης συστάδας αύξησε την ενδο-ομαδική διακύμανση (within-cluster sum of squares) χωρίς να προσφέρει διακριτό στρατηγικό πλεονέκτημα.

Σύμφωνα με τους **Wedel & Kannan (2016)**, η επιτυχία μιας τμηματοποίησης κρίνεται από το αν οι ομάδες είναι *Substantial* (αρκετά μεγάλες για να δικαιολογούν επένδυση) και *Actionable* (εφαρμόσιμες). Η δική μας επιλογή δημιουργεί τρεις πυλώνες με

απόλυτα διακριτά προφίλ Recency και Monetary, κάτι που στην εργασία των Lee & Jiang συχνά συγχεόταν, οδηγώντας σε αλληλεπικαλυπτόμενες στρατηγικές (cluster overlap).

4.6.3 Συμπεριφορική Σύγκλιση και το "RFM Paradox"

Μια σημαντική συμβολή της παρούσας εργασίας είναι η επιβεβαίωση αυτού που οι **Fader, Hardie & Lee (2005)** ονομάζουν "The RFM Paradigm". Στην ανάλυσή μας (Εικόνα 4.6), παρατηρήθηκε ότι το Cluster 1 (VIP) παρουσιάζει μια «αντιστρόφως ανάλογη» σχέση μεταξύ Recency και Frequency.

Σε σύγκριση με τις προγενέστερες αναλύσεις του **UCI Repository**, η έρευνά μας ανέδειξε ότι η μεταβλητή **Recency** δεν είναι απλώς ένας δείκτης χρόνου, αλλά ένας δείκτης «ψυχολογικής δέσμευσης». Όπως υποστηρίζουν οι **Lemon & Verhoef (2016)**, οι πελάτες με χαμηλό Recency είναι πιο δεκτικοί σε Cross-selling ενέργειες. Αυτό το εύρημα μας διαφοροποιεί από τις καθαρά τεχνικές προσεγγίσεις που αντιμετωπίζουν το RFM ως στατικό πίνακα, καθώς εμείς το αντιμετωπίζουμε ως **δυναμικό εργαλείο πρόβλεψης** της Δια Βίου Αξίας (CLV).

4.6.4 Γεωπολιτική Κατανομή και Ψηφιακό Οικοσύστημα

Ενώ η μελέτη των Chen et al. (2012) εστίασε στην εσωτερική αγορά του Ηνωμένου Βασιλείου, η δική μας χαρτογράφηση (Εικόνα 4.1) αποκάλυψε την ύπαρξη «δορυφορικών» VIP ομάδων στην Ευρωπαϊκή Κοινότητα. Αυτό το εύρημα έρχεται σε συμφωνία με τις τάσεις του "Cross-border E-commerce" που αναλύουν οι **Laudon & Traver (2022)**. Η εργασία μας αποδεικνύει ότι ο αλγόριθμος K-Means μπορεί να εντοπίσει ομοιογενή καταναλωτικά πρότυπα πέρα από εθνικά σύνορα, προσφέροντας στην επιχείρηση τη δυνατότητα για **Geographic Micro-targeting**.

Πίνακας 4.5: Ολιστική Σύγκριση Μεθοδολογικού και Θεωρητικού Πλαισίου

Μεθοδολογική Παράμετρος	Παραδοσιακή Προσέγγιση (Chen et al.)	Τεχνική Προσέγγιση (Lee & Jiang)	Παρούσα Μελέτη (Master's Thesis)	Ακαδημαϊκή Υπεροχή
Φιλοσοφία Δεδομένων	Data Cleaning (Deletion)	Data Transformation (Scaling)	Data Enrichment (Winsorization)	Προστασία του "Long Tail" της κερδοφορίας.
Κριτήριο Clustering	Elbow Method	Silhouette Score	Elbow & Managerial Heuristics	Εξασφάλιση επιχειρηματικής λειτουργικότητας.
Στρατηγική Εστίαση	Transactional (Συναλλακτική)	Algorithmic (Αλγοριθμική)	Relational (Σχεσιακή)	Σύνδεση με CLV και Retention Theory.
Ερμηνεία Recency	Χρονικό διάστημα	Στατιστική μεταβλητή	Δείκτης Δέσμευσης (Engagement)	Ευθυγράμμιση με Lemon & Verhoef (2016).

Πεδίο Εφαρμογής	Τοπική Αγορά	Γενικό Dataset	Διεθνές Ψηφιακό Μάρκετινγκ	Ανάδειξη ευκαιριών επέκτασης στην Ε.Ε.
-----------------	--------------	----------------	----------------------------	--

4.6.5 Τελική Σύνθεση και Προστιθέμενη Αξία

Η ανωτέρω συγκριτική ανάλυση τεκμηριώνει ότι η παρούσα διπλωματική εργασία δεν αποτελεί μια απλή αναπαραγωγή γνωστών μεθόδων, αλλά μια **κριτική εφαρμογή** τους. Η διαφοροποίησή μας στην προεπεξεργασία των δεδομένων και η εστίαση στην ερμηνευτική ισχύ των τριών συστάδων, προσφέρουν ένα μοντέλο που είναι ταυτόχρονα στατιστικά έγκυρο και διοικητικά αξιοποιήσιμο. Όπως καταλήγουν οι **Shmueli et al. (2017)**, η αξία των Business Analytics δεν έγκειται στην πολυπλοκότητα του αλγορίθμου, αλλά στην ικανότητά του να παράγει **αξιοποιήσιμη γνώση (Actionable Insights)**, στόχος που επετεύχθη πλήρως στην παρούσα μελέτη.

ΚΕΦΑΛΑΙΟ 5: ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΜΕΛΛΟΝΤΙΚΕΣ ΕΠΕΚΤΑΣΕΙΣ

5.1 Συμπεράσματα

Η παρούσα διπλωματική εργασία ολοκλήρωσε τη διερεύνηση της καταναλωτικής συμπεριφοράς στο περιβάλλον του ηλεκτρονικού εμπορίου, επιχειρώντας να γεφυρώσει το χάσμα ανάμεσα στη θεωρητική προσέγγιση του Στρατηγικού Μάρκετινγκ και την εφαρμοσμένη Επιστήμη των Δεδομένων. Η ανάλυση ανέδειξε ότι η μετάβαση από το παραδοσιακό, μαζικό μάρκετινγκ σε πρακτικές ακριβείας δεν αποτελεί πλέον επιλογή αλλά αναγκαιότητα για τη βιωσιμότητα των σύγχρονων επιχειρήσεων. Μέσα από τη μεθοδολογική σύζευξη του μοντέλου RFM με τον αλγόριθμο μηχανικής μάθησης K-Means, η έρευνα πέτυχε τον μετασχηματισμό ακατέργαστων δεδομένων συναλλαγών σε επιχειρησιακά αξιοποιήσιμη γνώση, επιβεβαιώνοντας πως τα δεδομένα αποτελούν το σημαντικότερο άυλο περιουσιακό στοιχείο της ψηφιακής οικονομίας.

Αναφορικά με την ποσοτικοποίηση της αγοραστικής συμπεριφοράς, η μελέτη κατέδειξε ότι η πολυδιάστατη προσέγγιση είναι σαφώς ανώτερη των μονοδιάστατων μετρήσεων. Η εφαρμογή του μοντέλου RFM, ενισχυμένη με τη διαδικασία της στατιστικής τυποποίησης μέσω των Z-Scores, αποδείχθηκε εξαιρετικά αποτελεσματική στην εξάλειψη των στρεβλώσεων που προκαλούν οι διαφορετικές κλίμακες μέτρησης μεταξύ της χρηματικής αξίας και της συχνότητας των αγορών. Η διαδικασία αυτή επέτρεψε την αντικειμενική αξιολόγηση της πελατειακής βάσης, αποκαλύπτοντας ότι η αγοραστική αξία δεν είναι μια στατική έννοια, αλλά μια σύνθετη μεταβλητή που διαμορφώνεται από την πρόσφατη αλληλεπίδραση, την επαναληψιμότητα και το οικονομικό αποτύπωμα του κάθε πελάτη.

Ένα από τα σημαντικότερα ευρήματα της έρευνας αφορά τη βέλτιστη τμηματοποίηση της αγοράς. Η ανάλυση ανέδειξε ότι ο διαχωρισμός της πελατειακής βάσης σε τρεις διακριτές συστάδες αποτελεί την πλέον αποδοτική λύση, καθώς εξισορροπεί ιδανικά τη στατιστική ακρίβεια με τη διοικητική ευχρηστία. Η επιβεβαίωση της Αρχής του Pareto, σύμφωνα με την οποία μια μικρή μειοψηφία πελατών (Cluster VIP) συνεισφέρει δυσανάλογα υψηλό μερίδιο στα συνολικά έσοδα, υπογραμμίζει την κρισιμότητα της αναγνώρισης και διατήρησης αυτού του τμήματος. Αντιθέτως, η ύπαρξη ενός μεγάλου πλήθους περιστασιακών πελατών καταδεικνύει τις λανθάνουσες ευκαιρίες για ανάπτυξη, αλλά και την ανάγκη για προσεκτική διαχείριση πόρων ώστε να μην σπαταλώνται ενέργειες μάρκετινγκ σε κοινό με χαμηλή πιθανότητα απόκρισης.

Σε στρατηγικό επίπεδο, τα αποτελέσματα υπαγορεύουν την υιοθέτηση διαφοροποιημένων προσεγγίσεων. Για τους πελάτες υψηλής αξίας απαιτείται εστίαση σε προγράμματα επιβράβευσης και προσωποποιημένης εξυπηρέτησης με στόχο τη μεγιστοποίηση της διάρκειας παραμονής τους στην επιχείρηση. Για τους πιστούς πελάτες, η στρατηγική οφείλει να επικεντρωθεί στην αύξηση της μέσης δαπάνης μέσω τεχνικών cross-selling, ενώ για τους περιστασιακούς αγοραστές ενδείκνυται η αυτοματοποιημένη επικοινωνία χαμηλού κόστους με στόχο την επαναδραστηριοποίηση. Συνεπώς, η παρούσα μελέτη παρέχει ένα ολοκληρωμένο πλαίσιο λήψης αποφάσεων, αποδεικνύοντας ότι η κατανόηση των διαφορετικών αναγκών των πελατών είναι το κλειδί για την απόκτηση ανταγωνιστικού πλεονεκτήματος.

5.2 Μελλοντικές Επεκτάσεις

Παρά τη σημαντική συνεισφορά της παρούσας εργασίας στην κατανόηση της πελατειακής συμπεριφοράς, η έρευνα υπόκειται σε συγκεκριμένους περιορισμούς που απορρέουν κυρίως από τη φύση των διαθέσιμων δεδομένων και τη στατική μεθοδολογική προσέγγιση. Οι περιορισμοί αυτοί, ωστόσο, δεν μειώνουν την αξία των ευρημάτων, αλλά αντιθέτως ανοίγουν νέους ορίζοντες για περαιτέρω επιστημονική διερεύνηση και εμπλουτισμό της μεθοδολογίας σε μελλοντικές μελέτες. Η απουσία δημογραφικών και ψυχογραφικών χαρακτηριστικών, καθώς και η εστίαση σε ένα συγκεκριμένο χρονικό παράθυρο, δημιουργούν την ανάγκη για πιο σύνθετα μοντέλα που θα μπορούν να ενσωματώσουν ποιοτικά χαρακτηριστικά και τη χρονική δυναμική.

Μια πρωταρχική κατεύθυνση για μελλοντική έρευνα αφορά τη μετάβαση από την περιγραφική στην προγνωστική αναλυτική. Ενώ η παρούσα μελέτη εστίασε στο τι συνέβη στο παρελθόν, μελλοντικές εργασίες μπορούν να αξιοποιήσουν αλγορίθμους Επιβλεπόμενης Μάθησης, όπως τα Τυχαία Δάση (Random Forests) ή τα Νευρωνικά Δίκτυα, για να προβλέψουν μελλοντικές συμπεριφορές. Συγκεκριμένα, η ανάπτυξη μοντέλων πρόβλεψης της πιθανότητας αποχώρησης (Churn Prediction) θα επέτρεπε στις επιχειρήσεις να παρεμβαίνουν προληπτικά πριν ο πελάτης διακόψει τη σχέση του με το κατάστημα. Παράλληλα, η ακριβέστερη εκτίμηση της μελλοντικής Δια Βίου Αξίας (Customer Lifetime Value) θα μπορούσε να βελτιστοποιήσει την κατανομή του προϋπολογισμού μάρκετινγκ, εστιάζοντας επενδύσεις σε πελάτες με τη μεγαλύτερη προοπτική απόδοσης.

Επιπλέον, η ενσωμάτωση τεχνικών Επεξεργασίας Φυσικής Γλώσσας (NLP) και Ανάλυσης Καλαθιού Αγορών θα μπορούσε να προσδώσει μια ποιοτική διάσταση στην ποσοτική ανάλυση. Η αξιοποίηση των κειμενικών περιγραφών των προϊόντων θα επέτρεπε την κατανόηση όχι μόνο του ύψους της δαπάνης, αλλά και των προτιμήσεων των πελατών σε επίπεδο κατηγορίας προϊόντων. Συνδυάζοντας αυτή την πληροφορία με την ανάλυση συσχέτισης, θα μπορούσαν να εντοπιστούν μοτίβα συνδυαστικών αγορών που χαρακτηρίζουν τις διαφορετικές συστάδες πελατών, οδηγώντας σε πιο στοχευμένες προτάσεις προϊόντων και βελτιώνοντας την εμπειρία του χρήστη.

Τέλος, κρίσιμη θεωρείται η μετάβαση από τη στατική στη δυναμική τμηματοποίηση. Η αγοραστική συμπεριφορά δεν είναι σταθερή, και οι πελάτες μετακινούνται συχνά μεταξύ των κατηγοριών. Η εφαρμογή ανάλυσης χρονοσειρών ή αλυσίδων Markov θα μπορούσε να χαρτογραφήσει την πορεία των πελατών στον χρόνο, εντοπίζοντας τα κρίσιμα σημεία καμπής όπου ένας πιστός πελάτης αρχίζει να εμφανίζει σημάδια αδράνειας. Μια τέτοια προσέγγιση θα καθιστούσε το σύστημα τμηματοποίησης έναν ζωντανό οργανισμό που προσαρμόζεται συνεχώς στις μεταβαλλόμενες συνθήκες της αγοράς, προσφέροντας στις επιχειρήσεις τη δυνατότητα να αντιδρούν άμεσα και αποτελεσματικά σε κάθε αλλαγή της καταναλωτικής συμπεριφοράς.

Βιβλιογραφία

- Aggarwal, C. C. (2015). *Data mining: The textbook*. Springer International Publishing.
- Agrawal, R., Imielinski, T., & Swami, A. (1993). Mining association rules between sets of items in large databases. In *Proceedings of the 1993 ACM SIGMOD international conference on Management of data* (pp. 207–216).
- Alpaydin, E. (2020). *Introduction to machine learning* (4th ed.). MIT Press.
- Anitha, P., & Patil, M. M. (2019). RFM model for customer segmentation. *International Journal of Recent Technology and Engineering*, 8(4), 2277-3878.
- Arthur, D., & Vassilvitskii, S. (2007). k-means++: The advantages of careful seeding. In *Proceedings of the eighteenth annual ACM-SIAM symposium on Discrete algorithms* (pp. 1027-1035).
- Ball, G. H., & Hall, D. J. (1965). *ISODATA, a novel method of data analysis and pattern classification*. Stanford Research Institute.
- Baturynska, I., & Martinsen, K. (2021). Optimization of manufacturing processes by customer segmentation and sales forecasting: A case study. *Applied Sciences*, 11(11), 1-18.
- Berry, M. J. A., & Linoff, G. S. (2004). *Data mining techniques: For marketing, sales, and customer relationship management*. Wiley Publishing.
- Bishop, C. M. (2006). *Pattern recognition and machine learning*. Springer.
- Box, G. E. P., & Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 26(2), 211-252.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5-32.
- Bryman, A. (2016). *Social research methods*. Oxford University Press.
- Buttle, F., & Maklan, S. (2019). *Customer relationship management: Concepts and technologies*. Routledge.
- Chaffey, D., & Ellis-Chadwick, F. (2019). *Digital marketing: Strategy, implementation and practice*. Pearson.

Καταναλωτικές Τάσεις στις Ηλεκτρονικές Αγορές μέσω Ανάλυσης Δεδομένων και Μοντέλων Μηχανικής Μάθησης

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(3), 197-208.

Cheng, C.-H., & Chen, M.-C. (2009). Classifying the segmentation of customer value via RFM model and RS theory. *Expert Systems with Applications*, 36(3), 4176-4184.

Christy, A. J., Umamakeswari, A., Priyatharsini, J., & Neyaa, A. (2018). RFM ranking – An effective approach to customer segmentation. *Journal of King Saud University - Computer and Information Sciences*, 33(10), 1-12.

Constantinides, E. (2004). Influencing the online consumer's behavior: The Web experience. *Internet Research*, 14(2), 111-126.

Davenport, T. H. (2006). Competing on analytics. *Harvard Business Review*, 84(1), 98-107.

Davies, D. L., & Bouldin, D. W. (1979). A cluster separation measure. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 1(2), 224-227.

Dua, D., & Graff, C. (2019). *UCI Machine Learning Repository*. University of California, Irvine, School of Information and Computer Sciences.

Dunn, J. C. (1973). A fuzzy relative of the ISODATA process and its use in detecting compact well-separated clusters. *Journal of Cybernetics*, 3(3), 32-57.

Ester, M., Kriegel, H.-P., Sander, J., & Xu, X. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *KDD-96 Proceedings* (pp. 226-231).

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. (1996). From data mining to knowledge discovery in databases. *AI Magazine*, 17(3), 37-54.

Few, S. (2012). *Show me the numbers: Designing tables and graphs to enlighten*. Analytics Press.

Forgy, E. W. (1965). Cluster analysis of multivariate data: efficiency versus interpretability of classifications. *Biometrics*, 21, 768-769.

Friedman, J. H. (2001). Greedy function approximation: A gradient boosting machine. *Annals of Statistics*, 29(5), 1189-1232.

Gower, J. C. (1971). A general coefficient of similarity and some of its properties. *Biometrics*, 27(4), 857-871.

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2018). *Multivariate data analysis*. Cengage Learning.

Han, J., Kamber, M., & Pei, J. (2012). *Data mining: Concepts and techniques*. Morgan Kaufmann.

Καταναλωτικές Τάσεις στις Ηλεκτρονικές Αγορές μέσω Ανάλυσης Δεδομένων και Μοντέλων Μηχανικής Μάθησης

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction*. Springer.

Huang, Z. (1998). Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data Mining and Knowledge Discovery*, 2(3), 283-304.

Hughes, A. M. (2005). *Strategic database marketing*. McGraw-Hill.

Jain, A. K. (2010). Data clustering: 50 years beyond K-means. *Pattern Recognition Letters*, 31(8), 651-666.

Jobber, D., & Ellis-Chadwick, F. (2019). *Principles and practice of marketing*. McGraw-Hill Education.

Kanungo, T., Mount, D. M., Netanyahu, N. S., Piatko, C. D., Silverman, R., & Wu, A. Y. (2002). An efficient k-means clustering algorithm: Analysis and implementation. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 24(7), 881-892.

Kaufman, L., & Rousseeuw, P. J. (2009). *Finding groups in data: An introduction to cluster analysis*. John Wiley & Sons.

Knaflic, C. N. (2015). *Storytelling with data: A data visualization guide for business professionals*. Wiley.

Kodinariya, T. M., & Makwana, P. R. (2013). Review on determining number of cluster in K-Means clustering. *International Journal of Advance Research in Computer Science and Management Studies*, 1(6), 90-95.

Kohonen, T. (1982). Self-organized formation of topologically correct feature maps. *Biological Cybernetics*, 43(1), 59-69.

Kotler, P., & Keller, K. L. (2016). *Marketing management* (15th ed.). Pearson.

Kuhn, M., & Johnson, K. (2013). *Applied predictive modeling*. Springer.

Kumar, V., & Reinartz, W. (2018). *Customer relationship management: Concept, strategy, and tools*. Springer.

Larose, D. T., & Larose, C. D. (2014). *Discovering knowledge in data: An introduction to data mining*. Wiley.

Laudon, K. C., & Traver, C. G. (2021). *E-commerce: Business, technology, society*. Pearson.

Linoff, G. S., & Berry, M. J. A. (2011). *Data mining techniques: For marketing, sales, and customer relationship management*. Wiley.

Lloyd, S. P. (1982). Least squares quantization in PCM. *IEEE Transactions on Information Theory*, 28(2), 129-137.

Καταναλωτικές Τάσεις στις Ηλεκτρονικές Αγορές μέσω Ανάλυσης Δεδομένων και Μοντέλων Μηχανικής Μάθησης

MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability* (Vol. 1, No. 14, pp. 281-297).

Milligan, G. W., & Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. *Psychometrika*, 50(2), 159-179.

Murtagh, F., & Contreras, P. (2012). Algorithms for hierarchical clustering: An overview. *WIREs Data Mining and Knowledge Discovery*, 2(1), 86-97.

Ngai, E. W. T., Xiu, L., & Chau, D. C. K. (2009). Application of data mining techniques in customer relationship management: A literature review and classification. *Expert Systems with Applications*, 36(2), 2592-2602.

Osborne, J. W. (2010). Improving your data transformations: Understanding Winsorization. *Practical Assessment, Research, and Evaluation*, 15(12), 1-10.

Peppers, D., & Rogers, M. (2016). *Managing customer experience and relationships: A strategic framework*. Wiley.

Provost, F., & Fawcett, T. (2013). *Data science for business: What you need to know about data mining and data-analytic thinking*. O'Reilly Media.

Quinlan, J. R. (1986). Induction of decision trees. *Machine Learning*, 1(1), 81-106.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, 20, 53-65.

Saunders, M., Lewis, P., & Thornhill, A. (2019). *Research methods for business students*. Pearson.

Sharda, R., Delen, D., & Turban, E. (2020). *Analytics, data science, & artificial intelligence: systems for decision support*. Pearson.

Steinley, D. (2006). K-means clustering: A half-century synthesis. *British Journal of Mathematical and Statistical Psychology*, 59(1), 1-34.

Tableau Software. (2024). *Visual analysis best practices* [Technical documentation]. Tableau Technical Publishing.

Tan, P.-N., Steinbach, M., & Kumar, V. (2018). *Introduction to data mining* (2nd ed.). Pearson.

Tibshirani, R., Walther, G., & Hastie, T. (2001). Estimating the number of clusters in a data set via the gap statistic. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 63(2), 411-423.

Tsiptsis, K., & Chorianopoulos, A. (2009). *Data mining techniques in CRM: Inside customer segmentation*. Wiley.

Tufte, E. R. (2001). *The visual display of quantitative information*. Graphics Press.

Vapnik, V. N. (1995). *The nature of statistical learning theory*. Springer.

Ward, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, 58(301), 236-244.

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), 4199-4206.

Witten, I. H., Frank, E., Hall, M. A., & Pal, C. J. (2016). *Data mining: Practical machine learning tools and techniques*. Morgan Kaufmann.

Wu, J. (2012). *Advances in K-means clustering: A data mining thinking*. Springer.

Zhang, T., Ramakrishnan, R., & Livny, M. (1996). BIRCH: An efficient data clustering method for very large databases. *SIGMOD Record*, 25(2), 103-114.