



ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ
ΣΧΟΛΗ ΔΙΟΙΚΗΤΙΚΩΝ, ΟΙΚΟΝΟΜΙΚΩΝ ΚΑΙ
ΚΟΙΝΩΝΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ
ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΔΙΟΙΚΗΣΗ ΕΠΙΧΕΙΡΗΣΕΩΝ – MBA
ΚΑΤΕΥΘΥΝΣΗ ΨΗΦΙΑΚΟ ΕΠΙΧΕΙΡΕΙΝ

Μεταπτυχιακή Διπλωματική Εργασία

Τίτλος Εργασίας

***Τμηματοποίηση Πελατών με Χρήση της Ανάλυσης RFM σε Εταιρεία Βελτίωσης Σπιτιού
& Ιδιοκατασκευής***

Συγγραφέας:

Κωνσταντίνος Ρήγας

ΑΜ:
MBA 20042

Επιβλέπων Καθηγητής
Δημήτριος Παπακυριακόπουλος

Αθήνα, Σεπτέμβριος 2022



**UNIVERSITY OF WEST ATTICA
SCHOOL OF MANAGEMENT, ECONOMICS AND
SOCIAL SCIENCES
DEPARTMENT OF BUSINESS ADMINISTRATION
POSTGRADUATE PROGRAM MBA
DIGITAL BUSINESS MAJOR**

Diploma Thesis

Title

Customer Segmentation of a Home Improvement & DIY Company, Using the RFM Analysis

Student name and surname:

Konstantinos Rigas

Registration number:

MBA 20042

Supervisor name and surname:

Dimitrios Papakyriakopoulos

Athens, September 2022



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ
ΣΧΟΛΗ ΔΙΟΙΚΗΤΙΚΩΝ, ΟΙΚΟΝΟΜΙΚΩΝ ΚΑΙ
ΚΟΙΝΩΝΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ
ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΔΙΟΙΚΗΣΗ ΕΠΙΧΕΙΡΗΣΕΩΝ – MBA
ΚΑΤΕΥΘΥΝΣΗ ΨΗΦΙΑΚΟ ΕΠΙΧΕΙΡΕΙΝ**

Τίτλος Εργασίας

Τμηματοποίηση Πελατών με Χρήση της Ανάλυσης RFM σε Εταιρεία Βελτίωσης Σπιτιού & Ιδιοκατασκευής

Μέλη Εξεταστικής Επιτροπής συμπεριλαμβανομένου και του Εισηγητή

Η μεταπτυχιακή διπλωματική εργασία εξετάστηκε επιτυχώς από την κάτωθι Εξεταστική Επιτροπή:

A/a	ΟΝΟΜΑ ΕΠΩΝΥΜΟ	ΒΑΘΜΙΑΔΑ/ΙΔΙΟΤΗΤΑ	ΨΗΦΙΑΚΗ ΥΠΟΓΡΑΦΗ
1	Δημήτριος Παπακυριακόπουλος	Επίκουρος καθηγητής / Επιβλέπων	
2	Ιωάννης Ψαρομήλιγκος	Καθηγητής / Μέλος	
3	Ιωάννης Ριζομυλιώτης	Επίκουρος καθηγητής / Μέλος	

ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΜΕΤΑΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ

Ο κάτωθι υπογεγραμμένος Κωνσταντίνος Ρήγας του Παναγιώτη, με αριθμό μητρώου MBA 20042, φοιτητής του Προγράμματος Μεταπτυχιακών Σπουδών MBA: DIGITAL BUSINESS του Τμήματος Διοίκησης Επιχειρήσεων της Σχολής Διοικητικών, Οικονομικών και Κοινωνικών Επιστημών του Πανεπιστημίου Δυτικής Αττικής, δηλώνω ότι:

«Είμαι συγγραφέας αυτής της μεταπτυχιακής εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της, είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, οι όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε ακριβώς είτε παραφρασμένες, αναφέρονται στο σύνολό τους, με πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο. Επίσης, βεβαιώνω ότι αυτή η εργασία έχει συγγραφεί από μένα αποκλειστικά και αποτελεί προϊόν πνευματικής ιδιοκτησίας τόσο δικής μου, όσο και του Ιδρύματος.

Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του πτυχίου μου».

Ο Δηλών

K. Rigas

*** Ονοματεπώνυμο /Ιδιότητα**

Δημήτριος Παπακυριακόπουλος / Επίκουρος Καθηγητής

Ψηφιακή Υπογραφή Επιβλέποντα

Ευχαριστίες

Έχοντας ταξιδέψει, εργαστεί και σπουδάσει σε διάφορες χώρες του εξωτερικού, κάθε ταξίδι ήταν διαφορετικό, και από το κάθε ταξίδι έχω αποκομίσει κάτι που το χρειάστηκα στη πορεία. Ένα “ταξίδι” που ήθελα να πραγματοποιήσω εδώ και πολύ καιρό, ήταν αυτό του μεταπτυχιακού. Ευτυχώς, μπόρεσα να πάρω μέρος και ήταν ένα από τα πιο ευχάριστα ταξίδια μέχρι τώρα. Είμαι χαρούμενος και περήφανος.

Θα ήθελα πρωτίστως να ευχαριστήσω θερμά τον επιβλέπων καθηγητή μου, τον Δρ. Δημήτρη Παπακυριακόπουλο, που με τις γνώσεις του και τον ενθουσιασμό του για την επιχειρηματική ευφυΐα, με ώθησε να οδεύσω σε αυτόν τον αρκετά δύσκολο αλλά συναρπαστικό δρόμο. Ελπίζω να τα καταφέρω, Δημήτρη. Σε ευχαριστώ για όλα, και να είσαι πάντα καλά.

Ακόμα ένα μεγάλο ευχαριστώ σε όλους τους καθηγητές του μεταπτυχιακού MBA – Digital Business, που μου έδωσαν χρήσιμα εφόδια για να τα καταφέρω στο ηλεκτρονικό μέλλον των επιχειρήσεων. Ήταν μια από τις καλύτερες επιλογές που έκανα για τις σπουδές μου.

Επίσης, ευχαριστώ την κοπέλα μου που είναι ακόμα δίπλα μου, ενώ έχω περάσει αμέτρητες ώρες προσπαθώντας να αποκτήσω αυτόν τον μεταπτυχιακό τίτλο. Σου χρωστάω.

Τέλος, ευχαριστώ την οικογένεια μου. Όσο περνάνε τα χρόνια, καταλαβαίνω καλύτερα την αξία όλων αυτών που μου έχετε δώσει, και το ποιος είμαι σήμερα. Αν και χιλιάδες χιλιόμετρα μακριά, πάντα ήσασταν και είστε δίπλα μου. Ελπίζω για πολλά χρόνια ακόμα.

The only thing you have to do in life,

Is to live.

Konstantinos Rigas

Περιεχόμενα

1	Εισαγωγή	3
1.1	Ερευνητικό Πρόβλημα & Στόχοι.....	3
1.2	Κίνητρα Εργασίας.....	7
1.3	Δομή εργασίας	8
2	Βιβλιογραφική Ανασκόπηση.....	10
2.1	Τμηματοποίηση Αγοράς	10
2.2	Τμηματοποίηση Πελατών	11
2.3	Χρήση της Τμηματοποίησης στο Μάρκετινγκ & στην Επιχειρησιακή Στρατηγική.....	12
2.4	Τύποι Τμηματοποίησης Πελατών.....	15
2.4.1	Με Βάση την Αξία (Value Based).....	15
2.4.2	Με βάση τη συμπεριφορά (Behavioral Based)	19
2.4.3	Με βάση τα δημογραφικά (Demographics Based).....	19
2.5	RFM Ανάλυση (Recency, Frequency, Monetary value)	21
2.5.1	Αξία ανάλυσης RFM.....	22
2.5.2	Χρήση ανάλυσης RFM	23
2.5.3	Μοντέλο ανάλυσης RFM.....	23
2.5.4	Παραδείγματα χρήσης της RFM στη πραγματική ζωή.....	26
3	Μεθοδολογία Έρευνας.....	33
3.1	Ερευνητική Μεθοδολογία	33
3.1.1	Design science – Επιστήμη του Σχεδιασμού	33
3.1.2	CRISP-DM - Εξόρυξη Δεδομένων.....	36
3.2	Διαθεσιμότητα Δεδομένων	42
3.2.1	Δεδομένα Πελατών.....	43
3.2.2	Δεδομένα Αγορών.....	51
3.2.3	Δεδομένα Προϊόντων.....	55
4	Εμπειρική Μελέτη	58
4.1	Πλαίσιο Λειτουργίας Επιχείρησης	58
4.2	Πειραματισμός	59
4.2.1	Παρουσίαση Συνόλου Δεδομένων Πελατών	59
4.2.2	Εκτέλεση Διαδικασίας RFM.....	60
4.2.3	Αποτελέσματα Διαδικασίας RFM	64
4.2.4	Ερμηνεία Αποτελεσμάτων RFM.....	70
4.2.5	Παρουσίαση Συνόλου Δεδομένων Προϊόντων	71
4.2.6	Εκτέλεση Διαδικασίας RFMV	71

4.2.7	Αποτελέσματα Διαδικασίας RFMV.....	75
4.2.8	Ερμηνεία Αποτελεσμάτων RFMV.....	84
4.2.9	Μελέτη Ακραίων Τιμών.....	85
5	Συζήτηση - Συμπεράσματα.....	89
5.1	Αποτελέσματα.....	89
5.2	Σύγκριση Αποτελεσμάτων με Άλλες Έρευνες.....	92
5.3	Περιορισμοί Έρευνας.....	93
5.4	Προτάσεις για Περαιτέρω Έρευνα.....	93
6	Βιβλιογραφία.....	95
7	Παράρτημα.....	99
7.1	Διαδικασία XML RFM.....	99
7.2	Διαδικασία XML RFMV.....	108

This page has been left intentionally blank.

Αυτή η σελίδα έχει παραμείνει σκόπιμα κενή.

Abstract

What is the value of a customer? Customers are essentially the lifeblood of a company, however, not everyone provides the same value to the business. There are customers that stopped buying, others that make rare visits to buy a few products and then, there are others that shop frequently and spend a significant amount on every purchase. For decades, marketers performed customer segmentation, helping them allocate resources more efficiently and achieve better results while minimizing costs. Customers can be segmented depending on their purchasing preferences (e.g. buying dog or cat products) or depending on how active are their purchasing habits during the last quarter. The problem that comes up with segmentation is that it is not a “one size fits all” technique. Different business needs require different approaches when it comes to customer segmentation. In order to tackle the complex and challenging problem of discovering the value of a company’s customers, the Recency – Frequency – Monetary Value (RFM) analysis provides a simple solution and a good base for further research.

This master’s thesis sets three research goals. The first goal is to study the value of a company’s customers, using the RFM analysis. Secondly, to review the RFM segmentation technique and its results, and last but not least, to improve upon the traditional analysis, and provide a new dimension. The new dimension will take into consideration the variety of products the customer has purchased when clustered with other customers.

For the aforementioned, a case study is conducted, using real purchasing data of 5.000 customers over the span of 2 years (2012 – 2014) of a US Home Improvement & DIY retail company, that belongs in Fortune 500. The research uses data mining techniques to cluster the customers based on their purchasing behaviors, as well as design science to implement the new dimension in the RFM analysis.

The results present 3 distinct clusters for both models (RFM & RFMV), ranging from inactive customers, to customers that have bought recently, buy frequently, spend significant amount of money and purchase a large variety of products.

All three goals have been achieved, as the RFM analysis has successfully clustered the customers based on the recency of their purchase, the frequency of their purchases and the monetary value spent, as well as the variety of products they purchased through the addition of the new dimension.

Proposals are provided to the company on how to adjust its marketing campaigns to each cluster for maximum efficiency and suggestions for further research are proposed.

Επιτελική Σύνοψη

Οι πελάτες είναι η πηγή ζωτικότητας της κάθε επιχείρησης. Όμως κάθε πελάτης έχει διαφορετική αξία για την επιχείρηση. Τα τμήματα μάρκετινγκ στην προσπάθεια τους να αυξήσουν την αποτελεσματικότητα των εκστρατειών τους, ομαδοποιούν πελάτες βάσει χαρακτηριστικών που μοιράζονται, και μια αξιόπιστη τεχνική τμηματοποίησης είναι η Recency – Frequency – Monetary Value (RFM). Η ανάλυση RFM ομαδοποιεί πελάτες βάσει της επικαιρότητας των αγορών τους, τη συχνότητα αυτών, αλλά και τη χρηματική αξία που δαπανούν στις αγορές τους. Βάσει της RFM, οι πελάτες που έχουν ψωνίσει πρόσφατα, αγοράζουν συχνά και κάνουν μεγάλες δαπάνες, έχουν μεγαλύτερη αξία για την επιχείρηση.

Η παρούσα διπλωματική θέτει τρεις ερευνητικούς στόχους. Ο πρώτος στόχος είναι η μελέτη της αξίας των πελατών μιας εταιρείας, χρησιμοποιώντας την ανάλυση RFM. Δεύτερον, να πραγματοποιήσει κριτική της τεχνικής τμηματοποίησης RFM αλλά και των αποτελεσμάτων της. Τέλος, να βελτιώσει την παραδοσιακή ανάλυση και να προσθέσει μια νέα διάσταση. Η νέα διάσταση θα λαμβάνει υπόψη την ποικιλία των προϊόντων που έχει αγοράσει ο πελάτης όταν ομαδοποιείται με άλλους πελάτες.

Για τα προαναφερθέντα, διεξάγεται μελέτη περίπτωσης, χρησιμοποιώντας πραγματικά αγοραστικά δεδομένα 5.000 πελατών σε διάστημα 2 ετών (2012 - 2014) μιας αμερικανικής εταιρείας λιανικής πώλησης βελτίωσης σπιτιού και ιδιοκατασκευής, η οποία ανήκει στη λίστα Fortune 500. Η έρευνα χρησιμοποιεί τεχνικές εξόρυξης δεδομένων για την ομαδοποίηση των πελατών με βάση την αγοραστική τους συμπεριφορά, καθώς και την επιστήμη του σχεδιασμού για την προσθήκη της νέας διάστασης στην ανάλυση RFM.

Τα αποτελέσματα παρουσιάζουν 3 διακριτές συστάδες και για τα δύο μοντέλα (RFM & RFMV), που κυμαίνονται από ανενεργούς πελάτες έως πελάτες που έχουν αγοράσει πρόσφατα, αγοράζουν συχνά, ξοδεύουν σημαντική ποσότητα χρημάτων και αγοράζουν μεγάλη ποικιλία προϊόντων.

Και οι τρεις στόχοι επιτεύχθηκαν, καθώς η ανάλυση RFM επιτυχώς ομαδοποίησε τους πελάτες με βάση την πιο πρόσφατη αγορά τους, τη συχνότητα των αγορών τους, τη χρηματική αξία που ξόδεψαν, καθώς και την ποικιλία των προϊόντων που αγόρασαν μέσω της προσθήκης της νέας διάστασης.

Παρέχονται προτάσεις προς την εταιρεία σχετικά με τον τρόπο προσαρμογής των εκστρατειών μάρκετινγκ σε κάθε συστάδα για μέγιστη αποτελεσματικότητα και προτείνονται προτάσεις για περαιτέρω έρευνα.

1 Εισαγωγή

Σε αυτό το κεφάλαιο, θα γίνει μια εισαγωγική παρουσίαση του προβλήματος που ασχολείται η παρούσα εργασία, μαζί με μια σύντομη εισαγωγή στην τμηματοποίηση πελατών, της ανάλυσης RFM και της αξίας τους.

1.1 Ερευνητικό Πρόβλημα & Στόχοι

Όπως έχουν διατυπώσει οι Gupta και Zeithaml (σελ. 718, 2006) “Οι πελάτες είναι η πηγή ζωτικότητας κάθε οργανισμού. Χωρίς πελάτες, μια επιχείρηση δεν έχει έσοδα, δεν έχει κέρδη και επομένως, δεν έχει αγοραία αξία.” Η ανάγκη κατανόησης της σχέσης μεταξύ της αξίας των πελατών και της κερδοφορίας της εταιρείας δεν ήταν ποτέ πιο κρίσιμη για την βιωσιμότητα και την εξέλιξη της επιχείρησης. Η μέτρηση της της Αξίας Χρόνου Ζωής των Πελατών (Customer Lifetime Value – CLV), και η παροχή των πόρων μάρκετινγκ στους πελάτες με την υψηλότερη δυνατή αξία πρέπει να βρίσκονται στον πυρήνα της επιχείρησης (Kumar et al. 2008). Ένας τρόπος για να επιτευχθεί είναι μέσω της συλλογής δεδομένων των πελατών κατά τις αγορές τους και την ανάλυση αυτών, ώστε να σκιαγραφηθεί το προφίλ τους και οι αγοραστικές τους προτιμήσεις.

Ο κάθε καταναλωτής έχει διαφοροποιημένες ανάγκες και προτιμήσεις (Kotler & Keller, 2016). Παίρνοντας για παράδειγμα τους πελάτες ενός καταστήματος με προϊόντα κατοικίδιων, ο πελάτης που έχει σκύλους θα αγοράσει προϊόντα για σκύλους, ενώ ο πελάτης που έχει γάτες, θα προμηθευτεί τα ανάλογα προϊόντα το δικό του κατοικίδιο.

Σύμφωνα με τον Davenport (2014), η αξιοποίηση των αγοραστικών δεδομένων των πελατών, μέσα από μια σειρά νέων τεχνολογικών πηγών δίνει στις εταιρείες μια πλουσιότερη κατανόηση της συμπεριφοράς και των προτιμήσεων των καταναλωτών - ανεξάρτητα από το αν οι καταναλωτές αυτοί είναι υφιστάμενοι ή μελλοντικοί πελάτες. Οι τεχνολογίες μεγάλων δεδομένων (Big Data) όχι μόνο κλιμακώνονται οικονομικότερα σε μεγαλύτερους όγκους δεδομένων, αλλά υποστηρίζουν και μια σειρά νέων τύπων δεδομένων και συσκευών. Η ευελιξία αυτών των τεχνολογιών είναι τόσο περιορισμένη όσο το όραμα του οργανισμού.

Πιο πρόσφατα, οι οργανισμοί μάρκετινγκ αλλά και τα τμήματα μάρκετινγκ των επιχειρήσεων αύξησαν ριζικά τους αναλυτικούς προσανατολισμούς τους με την άνοδο του λογισμικού διαχείρισης εκστρατειών. Τα στελέχη μάρκετινγκ που είναι προσανατολισμένα σε δεδομένα μπορούν πλέον να χρησιμοποιούν τα αποτελέσματα των αναλύσεων για να πειραματιστούν με διαφορετικές καμπάνιες για διαφορετικές ομάδες πελατών και να μάθουν ποιες καμπάνιες λειτουργούν καλύτερα σε ποια κοινά (Davenport, 2007). Πηγαίνοντας πίσω στο παράδειγμα με τους πελάτες που αγοράζουν προϊόντα κατοικίδιων, η εταιρεία μπορεί μέσω ανάλυσης των προτιμήσεων και των συμπεριφορών τους, να προωθήσει διαφορετικά μηνύματα μάρκετινγκ στον

καθένα. Ο πελάτης ο οποίος αγοράζει προϊόντα για γάτες θα λάβει ένα ηλεκτρονικό μήνυμα άμεσου μάρκετινγκ (Newsletter) με προϊόντα για γάτες και ενημερωτικό υλικό περί αυτό το κατοικίδιο, χωρίς να του εμφανίζονται προϊόντα ή πληροφορίες για άλλα κατοικίδια. Ο πελάτης που αγοράζει προϊόντα σκύλου θα λάβει με τη σειρά του ένα ενημερωτικό δελτίο για προϊόντα και πληροφορίες σκύλων.

Για να δημιουργηθεί ένα αποτελεσματικό μίγμα μάρκετινγκ, είναι σημαντικό να εντοπιστεί πρώτα μια συγκεκριμένη ομάδα πελατών που μοιράζονται παρόμοιες προτιμήσεις και ανταποκρίνονται σε συγκεκριμένα κανάλια επικοινωνίας και μηνύματα. Οι εφαρμογές τμηματοποίησης πελατών μπορούν να βοηθήσουν στον εντοπισμό διαφορετικών κοινοτήτων (τμημάτων) πελατών που μπορεί να μοιράζονται παρόμοια ενδιαφέροντα και συμπεριφορές (Fan et al. 2015). Η τμηματοποίηση πελατών πραγματοποιείται με βάση τα δεδομένα πελατών (δημογραφικά, αγοραστικές προτιμήσεις, κλπ.) σύμφωνα με τα οποία οι πελάτες μπορούν να ομαδοποιηθούν (Fotaki et al. 2014). Δεν υπάρχει κάποιο ενιαίο σχήμα τμηματοποίησης που να μπορεί να βοηθήσει αποτελεσματικά όλες τις επιχειρηματικές ανάγκες και τους στόχους. Υπάρχουν όμως διάφορες διαθέσιμες προσεγγίσεις τμηματοποίησης, καθεμία από τις οποίες είναι κατάλληλη για την υποστήριξη διαφορετικών επιχειρηματικών αναγκών (Tsiptsis & Chorianoopoulos, 2009).

Στο σημερινό έντονα ανταγωνιστικό επιχειρηματικό περιβάλλον, οι οργανισμοί αγωνίζονται να βελτιώσουν την πελατειακή εμπειρία, ώστε να επιτύχουν τη διατήρηση των πελατών που ήδη έχουν, και να αυξήσουν την πελατειακή τους βάση (Fotaki et al. 2014). Ο Massnick (1997) αναφέρει πως η βαθύτερη κατανόηση των πελατών επιβεβαιώνει την αξία της εστίασης σε αυτούς. Είναι εμπειρικά αποδεκτό ότι κοστίζει περίπου πέντε φορές περισσότερο μια επιχείρηση να κερδίσει έναν νέο πελάτη από ό,τι να διατηρήσει έναν υπάρχοντα, και δέκα φορές περισσότερο να επαναφέρει έναν δυσαρεστημένο πελάτη πίσω. Η ακαδημαϊκή κοινότητα της επιστήμης του μάρκετινγκ και τα στελέχη επιχειρήσεων αναγνωρίζουν την σημαντικότητα της διατήρησης των πελατών μιας εταιρείας. Όμως, σε μια επιχείρηση με χιλιάδες ή εκατομμύρια πελάτες, η πρόκληση που θα πρέπει να αντιμετωπιστεί αφορά την ποικιλία της πελατειακής βάσης. Στη βάση αναμένεται να υπάρχουν πελάτες που έχουν σταματήσει να αγοράζουν ή να εκδηλώνουν ενδιαφέρον για ένα και μόνο προϊόν, ορισμένοι άλλοι μπορεί να αγοράζουν πολύ συχνά, άλλοι να δαπανούν πολύ λίγο, ενώ κάποιοι άλλοι να πραγματοποιούν αγορές μεγάλης αξίας. Ο κάθε πελάτης προσδίδει διαφορετική αξία στην επιχείρηση και είναι σημαντικό ότι η αξία μπορεί να αναγνωριστεί σε διαφορετικές διαστάσεις όπως χρηματική αξία, όγκος συναλλαγών, συχνότητα επίσκεψης στο κατάστημα κλπ.

Το ερώτημα που πηγάζει από τα παραπάνω για μια εταιρεία είναι: **Ποια είναι η αξία των πελατών μιας επιχείρησης;** Η εταιρεία αρχικά θα πρέπει να αναλύσει τα δεδομένα των πελατών της, για να σκιαγραφήσει το προφίλ τους, ώστε να μπορεί να τους τμηματοποιήσει σε ομάδες με παρόμοια χαρακτηριστικά και να αναγνωρίσει την αξία τους και να στη συνέχεια να ενεργήσει με διαφορετικό τρόπο ανάλογα με το τμήμα. Με την επιτυχή τμηματοποίηση των πελατών της, μια

εταιρεία μπορεί εύκολα να προσαρμόσει τις ενέργειες μάρκετινγκ της για το κάθε τμήμα (συστάδα) πελατών βάσει των χαρακτηριστικών που μοιράζονται μεταξύ τους. Η παρούσα εργασία χρησιμοποιεί ισοδύναμα τις έννοιες της τμηματοποίησης και της συσταδοποίησης (clustering).

Στην εύρεση της αξίας των πελατών, πολλές επιχειρήσεις αξιοποιούν την ανάλυση Recency – Frequency – Monetary Value (RFM) είτε αποκλειστικά είτε συνδυαστικά με άλλες μεθόδους. Η RFM ανάλυση ομαδοποιεί τους πελάτες βάση 3 χαρακτηριστικών τους. Το πόσο πρόσφατα έχουν αγοράσει από την επιχείρηση ($R = \text{Recency} / \text{Επικαιρότητα}$), το πόσο συχνά κάνουν αγορές ($F = \text{Frequency} / \text{Συχνότητα}$) και τη χρηματική αξία των αγορών τους (Hughes, 2000). Οι πελάτες που αγοράζουν συχνά, έχουν πραγματοποιήσει πρόσφατα κάποια αγορά και δαπανούν μεγάλη ποσότητα χρημάτων, έχουν μεγαλύτερη αξία για την επιχείρηση από αυτούς που αγοράζουν σε πολύ αραιή συχνότητα, η τελευταία τους αγορά είναι αρκετά παλαιά, και τα χρήματα που αφήνουν με τις αγορές τους στην επιχείρηση είναι λιγιστά.

Για τις ανάγκες αυτή της έρευνας, ακολουθείται η ερευνητική σχεδίαση της μελέτης περίπτωσης (case study). Ειδικότερα, εφαρμόζεται η ανάλυση RFM σε πραγματικά δεδομένα λιανικής εταιρείας βελτίωσης σπιτιού και ιδιοκατασκευής, για να βρεθεί η αξία αυτών των πελατών. Τα δεδομένα που χρησιμοποιούνται είναι οι συναλλαγές πέντε χιλιάδων νοικοκυριών σε διάστημα 2 ετών. Η εταιρεία της έρευνας ανήκει στις 500 μεγαλύτερες επιχειρήσεις του κόσμου (Fortune 500), και επιθυμεί να διατηρήσει την ανωνυμία της.

Σε αυτή την έρευνα, τίθενται τρεις ερευνητικοί στόχοι:

1. Να μελετηθεί η αξία των πελατών της εταιρείας με χρήση της ανάλυσης RFM

Σε μια μεγάλη εταιρεία με εκατομμύρια πελάτες, είναι σημαντικό να προσδιορισθεί η αξία αυτών. Η εύρεση των διάφορων ομάδων πελατών με τις ξεχωριστές τους αξίες αποτελεί κλειδί για την επιχείρηση επειδή μπορεί να δημιουργήσει καλύτερα στοχευμένες καμπάνιες μάρκετινγκ, να κατανέμει καλύτερα τους πόρους της αλλά και να εξοικονομήσει χρήματα σε ενέργειες που δε θα αποφέρουν το επιθυμητό αποτέλεσμα. Είναι ενδιαφέρον να τονισθεί ότι η αξία του πελάτη είναι ένα δυναμικό μέγεθος μιας και μεταβάλλεται στο χρόνο και δεν υπάρχει μια αντικειμενική τεχνική ευρέως αποδεκτή που με ακρίβεια να μετράει το φαινόμενο. Η συνεισφορά της εργασίας είναι ότι προσφέρει μια οπτική της RFM από την πλευρά της επιχειρηματικής ευφυΐας (Business Intelligence).

2. Να γίνει κριτική της τεχνικής τμηματοποίησης RFM και των αποτελεσμάτων που δημιουργεί

Όπως προαναφέρθηκε, η αξία τμηματοποίησης των πελατών έχει μεγάλη αξία για την επιχείρηση. Η ανάλυση RFM αποτελεί μια παραδοσιακή ανάλυση τμηματοποίησης των πελατών και ανάδειξης της αξίας τους καθώς τους συσταδοποιεί βάση τριών πολύ κρίσιμων διαστάσεων. Πόσο πρόσφατα ήταν η τελευταία τους αγορά, πόσο συχνά κάνουν αγορές και ποια η χρηματικά αξία των αγορών τους. Η ανάλυση RFM μπορεί να εμφανίσει τις διάφορες ομάδες πελατών με τις ξεχωριστές τους αξίες. Πολλές μελέτες έχουν εφαρμόσει την τεχνική τμηματοποίησης RFM, με τα αποτελέσματα να εμφανίζουν επιτυχώς την αξία των πελατών για τη κάθε επιχείρηση ή οργανισμό που εφαρμόστηκε αυτή η τεχνική. Τμηματοποιώντας πελάτες μέσω της ανάλυσης RFM έχει ενδιαφέρον γιατί είναι μια απλή τεχνική που αντιμετωπίζει ένα αρκετά σύνθετο και δύσκολο πρόβλημα. Η θέση της εργασίας είναι ότι μια απλή τμηματοποίηση RFM μπορεί να δώσει μια πραγματική και επίκαιρη εικόνα για τις 3 διαστάσεις που αναλύει, θέτοντας μια καλή αρχή για περαιτέρω μελέτη.

3. Να ενισχυθεί η τεχνική τμηματοποίησης RFM με την προσθήκη καινούργιας διάστασης

Σε προηγούμενες έρευνες, όπως για παράδειγμα των Zhang et al. (2014) και Heldt et al. (2021), έχουν προταθεί επιπλέον διαστάσεις στην ανάλυση RFM. Οι Zhang et al. (2014), πρόσθεσαν τη διάσταση “C” (Clumpiness), όπου παίρνει υπόψη την συσσωμάτωση που δημιουργείται ανάμεσα στους πελάτες σε διάφορα χαρακτηριστικά όπως τα δημογραφικά και σε τι ενέργειες μάρκετινγκ ανταποκρίνονται. Με την προσθήκη αυτή της διάστασης, όχι μόνο η επιχείρηση έχει καλύτερη εικόνα του προφίλ των πελατών της, αλλά επιπλέον, οι πελάτες που αποτελούν τη κάθε συστάδα έχουν μεγαλύτερη ομοιομορφία μεταξύ τους. Σε αυτή την έρευνα, θα γίνει προσπάθεια μελέτης μιας καινούργιας διάστασης ώστε να προστεθεί επιπλέον αξία στην ανάλυση RFM και να σκιαγραφηθεί αποτελεσματικότερα το προφίλ των πελατών της εταιρείας της έρευνας. Η επέκταση που επιλέχθηκε είναι αυτή της “Ποικιλίας Προϊόντων” ή αλλιώς “V” (Variety). Η διαδικασία συσταδοποίησης θα παίρνει υπόψη και πόσα διαφορετικά προϊόντα έχει αγοράσει ο πελάτης μαζί με τις υπόλοιπες διαστάσεις της RFM.

1.2 Κίνητρα Εργασίας

Είναι γνωστό από προηγούμενες έρευνες όπως του Hughes (2000), των Kotler και Keller (2016) και της Fotaki et al. (2014), το πόσο σημαντική και ωφέλιμη είναι η τμηματοποίηση πελατών για μια επιχείρηση. Η τμηματοποίηση πελατών δεν ωφελεί μόνο τα στελέχη μάρκετινγκ αλλά και διάφορα άλλα άτομα με ποικίλους ρόλους που μπορούν να χρησιμοποιήσουν τα αποτελέσματα μιας τέτοιας ανάλυσης. Η παρούσα έρευνα στηρίζεται σε πραγματικά δεδομένα μίας μεγάλης επιχείρησης, όμως η ανάλυση μπορεί να χρησιμοποιηθεί σε οποιαδήποτε εταιρεία που μπορεί να παρέχει τα απαιτούμενα αγοραστικά δεδομένα των πελατών της.

Για αυτό το λόγο, κύριο κίνητρο για αυτή την έρευνα είναι η ανάδειξη της σημαντικότητας πρωτίστως της τμηματοποίησης πελατών, αλλά και της τεχνικής RFM. Η σημασία της τμηματοποίησης έχει τονιστεί αρκετά τις τελευταίες δεκαετίες, μέσα από έρευνες όπως αυτή του Kottler (2003), του Dwyer (1997), της Fotaki et al. (2014) και άλλων. Μία τεχνική τμηματοποίησης που πραγματοποιείται σε έρευνες είναι η RFM, όπου μέσα από μελέτες όπως αυτή του Hughes (2005), του Zhang et al. (2014) και του Lewis et al. (2006) έχουν αναδειχθεί τα οφέλη της. Η RFM δίνει μεγαλύτερη ευελιξία σε επικοινωνίες μάρκετινγκ καθώς τα στελέχη μάρκετινγκ είναι σε θέση να πραγματοποιήσουν διαφορετικές ενέργειες στα διαφορετικά τμήματα, αποφέροντας καλύτερα αποτελέσματα εκστρατειών μάρκετινγκ.

Μια τμηματοποίηση πελατών με χρήση τη τεχνική RFM είναι ένας πολύ καλός θεμελιώδης πυλώνας όπου μετέπειτες έρευνες μπορούν να στηριχθούν. Σε αυτή την έρευνα θα γίνει προσπάθεια επέκτασης της τεχνικής, προσθέτοντας τη διάσταση της ποικιλίας προϊόντων που αγοράζουν οι πελάτες. Μέσα από αυτή τη προσθήκη, τα αποτελέσματα θα φανούν χρήσιμα και ενδιαφέροντα όχι μόνο σε υπεύθυνους μάρκετινγκ για τη καλύτερη προώθηση προϊόντων, αλλά και σε υπευθύνους διαχείρισης των προσφερόμενων προϊόντων ώστε να παρέχονται ακόμα περισσότερα σε πελάτες που αγοράζουν μεγάλη ποικιλία. Τα οφέλη όμως δε σταματάνε εκεί. Το παραδοσιακό μοντέλο RFM αφήνει άνετο χώρο για διάφορες επεκτάσεις, όπου ένας αναλυτής μπορεί να αντλήσει επιπλέον πληροφορίες και να το προσαρμόσει σε κάθε περίπτωση διαφορετικά. Μέσω της RFM, οι ενέργειες μάρκετινγκ γίνονται πιο ευέλικτες και αποτελεσματικές, αλλά και οι εταιρικές αποφάσεις είναι πλέον καλύτερα ενημερωμένες γνωρίζοντας τις διαφορετικές προτιμήσεις των τμηματοποιημένων πελατών. Συγκεκριμένα, αποφάσεις όπως σε ποιο τμήμα πελατών θα εστιάσει η επιχείρηση (π.χ. επανενεργοποίηση παλιών πελατών, προσέλκυση νέων, αύξηση αξίας πιστών πελατών, κλπ.) ή αν θα πρέπει να κάνει αλλαγές στη διάθεση των προϊόντων της, μπορούν να βασιστούν στα αποτελέσματα της ανάλυσης RFM, αλλά και της επέκτασης που προτείνεται σε αυτή την εργασία.

Η παρούσα μελέτη μπορεί επίσης να φανεί ωφέλιμη και στο ακαδημαϊκό κοινό. Πολλές έρευνες έχουν χρησιμοποιήσει την ανάλυση RFM, και μεγάλη έμφαση έχει δοθεί στις επεκτάσεις που μπορούν να προστεθούν στο βασικό μοντέλο. Η προτεινόμενη επέκταση “V” στην RFM ελπίζει

να παρέχει ωφέλιμα αποτελέσματα για την επιχείρηση. Η ιδιαίτερη φύση του κλάδου DIY έχει μελετηθεί εκτενώς και ένα ενδιαφέρον χαρακτηριστικό είναι ότι οι πελάτες ορισμένες φορές καλούνται prosumers γιατί παράγουν (produce) και καταναλώνουν ταυτόχρονα (consumer) (Kotler, 2010). Η πρόταση του Kotler (2010) αναγνωρίζει δυο βασικές ομάδες (τμήματα) prosumers και εννοιολογικά αναγνωρίζονται ως Avid Hobbyist και Archprosumer. Ενώ διαθέτουν διαφορετικά κίνητρα παραγωγής-κατανάλωσης ωστόσο και στα δυο υπάρχει η υπόθεση ότι συνεχώς παράγουν και μάλιστα διαφορετικά προϊόντα κάθε φορά. Συνεπώς η έννοια του prosumer περιέχει την συνεχή προμήθεια διαφορετικών προϊόντων και αυτή είναι η ιδέα όπου στηρίζεται η εργασία προκειμένου να επεκτείνει την παραδοσιακή RFM ανάλυση.

1.3 Δομή εργασίας

Η παρούσα έρευνα χωρίζεται σε 7 κεφάλαια, συμπεριλαμβανομένης της εισαγωγής.

Στο 2^ο κεφάλαιο, γίνεται ανασκόπηση της σχετικής βιβλιογραφίας. Συγκεκριμένα, θα αναλυθεί η τμηματοποίηση αγοράς και τμηματοποίηση πελατών μέσα από προηγούμενες έρευνες. Επίσης, θα παρουσιαστούν οι χρήσεις της τμηματοποίησης στο μάρκετινγκ, μερικοί διαφορετικοί τύποι τμηματοποίησης και τέλος, θα γίνει ανασκόπηση της βιβλιογραφίας για την ανάλυση RFM, και ειδικότερα, της αξίας της, των χρήσεων της, το μοντέλο εφαρμογής που χρησιμοποιούν συχνά οι ερευνητές αλλά και μερικά παραδείγματα χρήσης της RFM στην πραγματική ζωή με τα αποτελέσματά τους.

Στο 3^ο κεφάλαιο θα παρουσιαστεί η μεθοδολογία της έρευνας. Αρχικά, θα παρουσιαστεί η Επιστήμη του Σχεδιασμού και η καινοτομία αυτής της έρευνας, και δεύτερον θα παρουσιαστεί η μεθοδολογία Εξόρυξης Δεδομένων. Τέλος, θα γίνει παρουσίαση των δεδομένων που έχουν χορηγηθεί από την εταιρεία.

Στο 4^ο κεφάλαιο, αυτό της Εμπειρικής Μελέτης, θα παρουσιαστεί το πλαίσιο λειτουργίας της επιχείρησης και των καταναλωτών της, και στη συνέχεια, θα παρουσιαστεί διεξοδικά ο πειραματισμός της έρευνας. Ειδικότερα, θα γίνει παρουσίαση των δεδομένων των πελατών και των προϊόντων που αγόρασαν, θα παρουσιαστεί η διαδικασία της τμηματοποίησης με χρήση της RFM αλλά και της RFMV που είναι η καινούργια διάσταση, και τέλος θα γίνει παρουσίαση των αποτελεσμάτων και προσπάθεια ερμηνείας αυτών.

Στο 5^ο κεφάλαιο θα γίνει συνοψισμός των αποτελεσμάτων της εμπειρικής μελέτης, θα γίνει παρουσίαση των αποτελεσμάτων των στόχων που είχαν τεθεί στο πρώτο κεφάλαιο, θα παρουσιαστούν τα κύρια συμπεράσματα από την δημιουργία της τμηματοποίησης των πελατών του δείγματος, θα προσφερθούν προτάσεις για τη κάθε συστάδα και θα γίνει προσπάθεια σύγκρισης με άλλες μελέτες. Τέλος, θα αναφερθούν περιορισμοί που υπήρξαν στην έρευνα και προτάσεις για μελλοντική μελέτη.

Στο 6^ο κεφάλαιο θα παρατεθεί η βιβλιογραφία που χρησιμοποιήθηκε για αυτή την έρευνα, και τέλος, στο 7^ο κεφάλαιο θα παρατεθούν οι διαδικασίες που χρησιμοποιήθηκαν για την εμπειρική μελέτη σε γλώσσα XML.

2 Βιβλιογραφική Ανασκόπηση

Σε αυτό το κεφάλαιο, γίνεται ανασκόπηση της βιβλιογραφίας. Συγκεκριμένα, θα αναλυθεί η τμηματοποίηση αγοράς και τμηματοποίηση πελατών μέσα από προηγούμενες έρευνες. Επίσης, θα παρουσιαστούν οι χρήσεις της τμηματοποίησης στο μάρκετινγκ, μερικοί διαφορετικοί τύποι τμηματοποίησης και τέλος, θα γίνει ανασκόπηση της βιβλιογραφίας για την ανάλυση RFM, και ειδικότερα, της αξίας της, των χρήσεων της, το μοντέλο εφαρμογής που χρησιμοποιούν συχνά οι ερευνητές αλλά και μερικά παραδείγματα χρήσης της RFM στην πραγματική ζωή με τα αποτελέσματά τους.

2.1 Τμηματοποίηση Αγοράς

Η έννοια της τμηματοποίησης της αγοράς εισάγεται από τον Smith (1956) όπου αναφέρει πως «η τμηματοποίηση της αγοράς περιλαμβάνει τη θεώρηση μιας ετερογενούς αγοράς ως ένα σύνολο μικρότερων ομοιογενών αγορών, ως απάντηση στις διαφορετικές προτιμήσεις, οι οποίες οφείλονται στις επιθυμίες των καταναλωτών για ακριβέστερη ικανοποίηση των διαφορετικών επιθυμιών τους.»

Ο Smith προτείνει τρία κριτήρια που πρέπει να πληρούνται στην τμηματοποίηση:

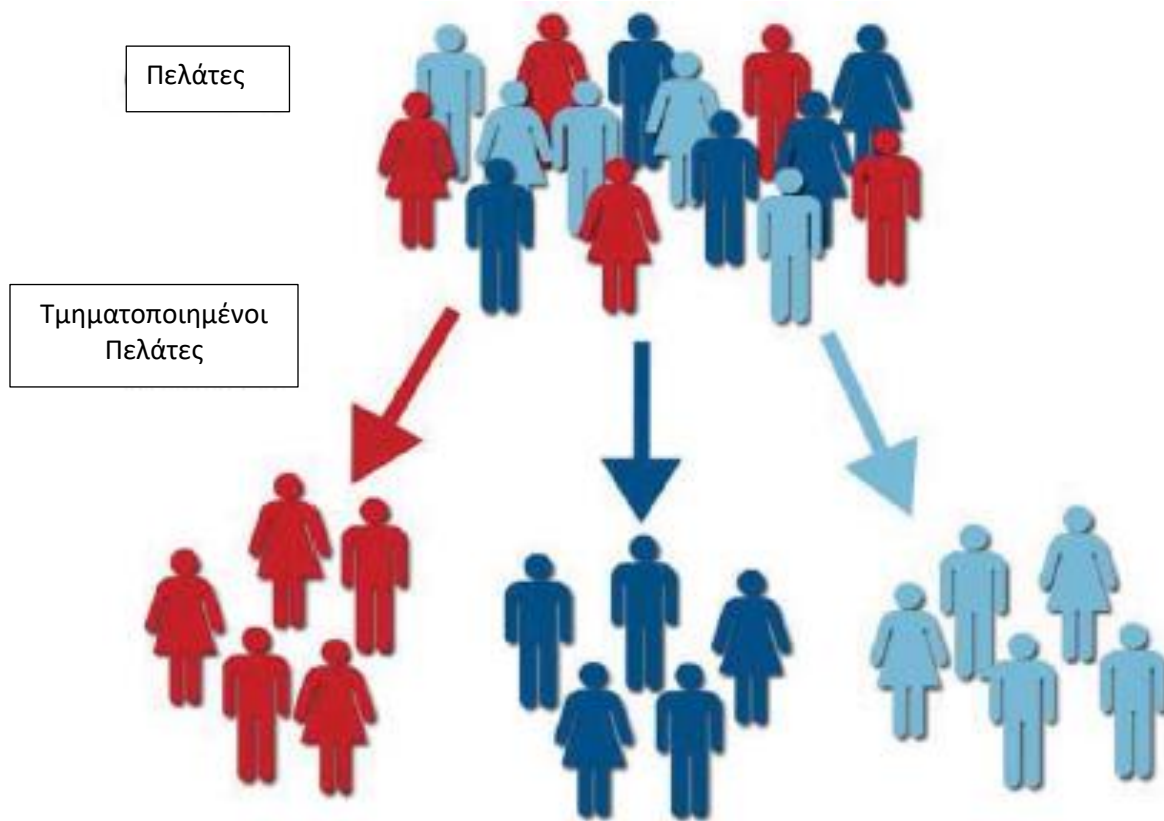
1. Ομοιογένεια (δηλ. ομοιότητα αναγκών εντός της ομάδας),
2. Διάκριση (δηλ. μοναδικότητα μεταξύ των ομάδων) και
3. Αντίδραση (δηλ. ομοιότητα της αντίδρασης απέναντι στη στρατηγική μάρκετινγκ, το προϊόν, την προσφορά ή τις υπηρεσίες εντός της ομάδας).

Ο Kottler (2003) ισχυρίζεται ότι η τμηματοποίηση της αγοράς επιτρέπει τη δημιουργία ενός πιο λεπτομερούς προϊόντος ή υπηρεσίας που προσφέρει και την κατάλληλη τιμή για ένα τμήμα-στόχο. Ισχυρίζεται επίσης ότι τα στελέχη μάρκετινγκ μπορούν να παρέχουν καλύτερα κανάλια διανομής και επικοινωνίας στο τμήμα-στόχο.

Συνήθως χρησιμοποιούνται τρεις κύριες διαστάσεις τμηματοποίησης: η δημογραφική διάσταση, η ψυχογραφική και η συμπεριφορική. Οι δημογραφικές μεταβλητές είναι οι πιο δημοφιλείς και περιλαμβάνουν την ηλικία, το μέγεθος της οικογένειας, τον κύκλο ζωής της οικογένειας, το φύλο, το εισόδημα, το επάγγελμα, την εκπαίδευση, την θρησκεία, τη φυλή, την εθνικότητα και την κοινωνική τάξη (Kottler, 2003). Ο σκοπός των ψυχογραφικών στοιχείων είναι η απόκτηση μιας καλύτερης κατανόησης του καταναλωτή ως ατόμου μέσω της μέτρησης των ψυχολογικών διαστάσεων, του τρόπου ζωής, των ενδιαφερόντων και των απόψεών του (Ziph, 1971). Η τμηματοποίηση της συμπεριφοράς επικεντρώνεται στην πραγματική συμπεριφορά των χρηστών, συμπεριλαμβανομένων των περιστάσεων, των οφελών, της κατάστασης του χρήστη, του ποσοστού χρήσης, της κατάστασης αφοσίωσης, της ετοιμότητας και της στάσης απέναντι στα προϊόντα (Rokeach, 1973).

2.2 Τμηματοποίηση Πελατών

Η τμηματοποίηση πελατών είναι η διαδικασία διαχωρισμού των πελατών σε διακριτές, ουσιαστικές και ομοιογενείς υποομάδες με βάση διάφορα χαρακτηριστικά και ιδιότητες. Χρησιμοποιείται ως εργαλείο διαφοροποίησης για το μάρκετινγκ. Επιτρέπει στους οργανισμούς να κατανοήσουν τους πελάτες τους και να αναπτύξουν διαφοροποιημένες στρατηγικές, προσαρμοσμένες στα χαρακτηριστικά τους (Tsiptsis & Chorianoopoulos, 2009).



Εικόνα 1: Τμηματοποίηση Πελατών (Petersen, 2019)

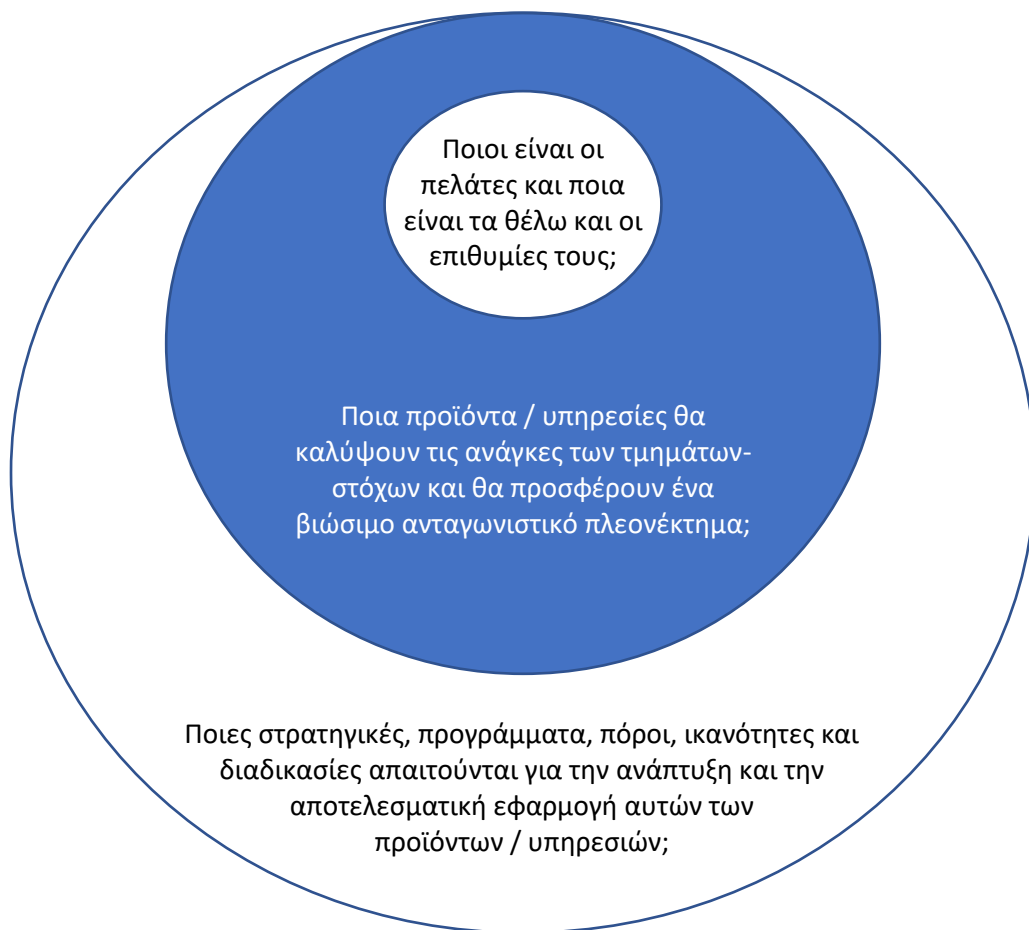
Η τμηματοποίηση, όπως απεικονίζεται και στην εικόνα 1, είναι μια θεμελιώδης στρατηγική για τη διαχείριση των προσπαθειών μάρκετινγκ που απευθύνονται στους πελάτες. Εισήχθη τη δεκαετία του 1950 για να αντικατοπτρίσει μια αλλαγή από το μαζικό μάρκετινγκ στη νέα (τότε) έννοια του μάρκετινγκ: τη στόχευση προϊόντων ή/και εκστρατειών μάρκετινγκ σε συγκεκριμένες ομάδες πελατών (Lemon & Mark, 2008).

Η τμηματοποίηση πελατών έχει ουσιαστικά απεριόριστες δυνατότητες ως εργαλείο που μπορεί να καθοδηγήσει τις επιχειρήσεις προς πιο αποτελεσματικούς τρόπους προώθησης των προϊόντων και να αναπτύξουν νέα προϊόντα (Cooil et al. 2008).

Η έννοια της τμηματοποίησης είναι κεντρική για το μάρκετινγκ. Στις αρχές εφαρμογής του στο μάρκετινγκ, η διαδικασία διαίρεσης ενός πληθυσμού πελατών μέσω τεχνικών ομαδοποίησης σε ομοιογενείς ομάδες γινόταν συχνά χωρίς τη χρήση μιας εξαρτημένης μεταβλητής/στόχου. Ωστόσο, τα στελέχη μάρκετινγκ συνειδητοποίησαν ότι η τμηματοποίηση δεν πρέπει να είναι αυτοσκοπός, αλλά μάλλον ένα μέσο για την επίτευξη ενός σκοπού. Καθώς οι περισσότερες εταιρείες θέλουν να μεγιστοποιήσουν τα κέρδη (ή κάποιο άλλο στόχο, π.χ. πωλήσεις), τα στελέχη μάρκετινγκ γρήγορα συνειδητοποίησαν ότι μια τμηματοποίηση θα πρέπει να διασφαλίζει ότι οι "καλύτεροι" πελάτες πρέπει να διαχωρίζονται από τους υπόλοιπους πελάτες (Jonker et al. 2004).

2.3 Χρήση της Τμηματοποίησης στο Μάρκετινγκ & στην Επιχειρησιακή Στρατηγική

Εννοιολογικά, κάθε επιχειρηματική στρατηγική θα πρέπει να βασίζεται στην κατανόηση, την ικανοποίηση, ή ακόμη και την υπέρβαση των αναγκών των τμημάτων-στόχων. Το σχήμα 1 αναδεικνύει τον κεντρικό ρόλο της τμηματοποίησης και την εξέλιξη των θεμελιωδών ερωτημάτων που πρέπει να αντιμετωπιστούν. Στο επίκεντρο βρίσκεται ο προσδιορισμός της υφιστάμενης και δυνητικής πελατειακής βάσης, η κατανόηση της υποκείμενης ετερογένειας και των εξελισσόμενων αναγκών και επιθυμιών των τμημάτων-στόχων. Ακολουθεί η ανταπόκριση στην τμηματοποίηση, δηλαδή οι κατευθυντήριες γραμμές για την ανάπτυξη προϊόντων και υπηρεσιών, καθώς και η σχετική τοποθέτησή τους για την κάλυψη των εξελισσόμενων αναγκών των τμημάτων-στόχων. Τέλος, η τοποθέτηση του προϊόντος αποτελεί τη βάση για την υπόλοιπη στρατηγική μάρκετινγκ και τις διαδικασίες, τις αποφάσεις κατανομής πόρων και άλλες δραστηριότητες της επιχείρησης (Baker & Hart, 2008).



Σχήμα 1: Στρατηγική με γνώμονα την αγορά (Baker & Hart, 2008)

Παραδοσιακά οι οργανισμοί, ανεξάρτητα από τον κλάδο στον οποίο δραστηριοποιούνται, τείνουν να χρησιμοποιούν συστήματα τμηματοποίησης της αγοράς που βασίζονται σε δημογραφικά στοιχεία και πληροφορίες σχετικά με την αξία. Τις τελευταίες δεκαετίες οι οργανισμοί αποφασίζουν τις δραστηριότητες μάρκετινγκ τους και αναπτύσσουν τα νέα προϊόντα και τις υπηρεσίες τους με βάση αυτά τα απλά, επιχειρηματικού κανόνα τμήματα (Ziafat & Shakeri, 2014).

Στις σημερινές ανταγωνιστικές αγορές, η προσέγγιση αυτή δεν είναι επαρκής. Αντιθέτως, οι οργανισμοί πρέπει να έχουν μια πλήρη εικόνα των πελατών τους προκειμένου να αποκτήσουν ανταγωνιστικό πλεονέκτημα. Πρέπει επίσης να εστιάζουν στις ανάγκες, τις επιθυμίες, τις στάσεις, τις συμπεριφορές, τις προτιμήσεις και τις αντιλήψεις των πελατών τους και να αναλύουν τα σχετικά δεδομένα για να προσδιορίσουν τα υποκείμενα τμήματα. Ο εντοπισμός ομάδων με μοναδικά χαρακτηριστικά θα επιτρέψει στον οργανισμό να τους διαχειριστεί και να τους στοχεύσει πιο αποτελεσματικά, μεταξύ άλλων, με εξατομικευμένες προσφορές προϊόντων και προωθητικές ενέργειες (Tsiptsis & Choriantopoulos, 2009).

Σύμφωνα με τους Lamb et al. (2011), το πρώτο βήμα για την τμηματοποίηση της αγοράς είναι η "επιλογή μιας αγοράς ή μιας κατηγορίας προϊόντων για μελέτη". Μπορεί να είναι μια αγορά την οποία η επιχείρηση έχει ήδη καταλάβει ή μια εντελώς νέα αγορά.

Το δεύτερο βήμα είναι η "επιλογή μιας βάσης ή βάσεων για την τμηματοποίηση της αγοράς". Αυτό το βήμα απαιτεί διοικητική διορατικότητα, δημιουργικότητα και γνώση της αγοράς. Δεν υπάρχουν επιστημονικές διαδικασίες για την επιλογή των μεταβλητών τμηματοποίησης. Ωστόσο, ένα επιτυχημένο σχέδιο τμηματοποίησης πρέπει να παράγει τμήματα της αγοράς που να πληρούν τα τέσσερα βασικά κριτήρια: "Ουσιαστικότητα, αναγνωρισιμότητα, προσβασιμότητα και ευθυγράμμιση" (Sun, 2009).

Το τρίτο βήμα είναι η "επιλογή περιγραφικών παραμέτρων τμηματοποίησης". Αφού η εταιρεία επιλέξει μία ή περισσότερες βάσεις, το στέλεχος μάρκετινγκ πρέπει να επιλέξει τις παραμέτρους τμηματοποίησης. Οι παράμετροι προσδιορίζουν τις συγκεκριμένες μεταβλητές τμηματοποίησης που θα χρησιμοποιηθούν.

Το τέταρτο είναι η "σκιαγράφηση και ανάλυση των τμημάτων". Η ανάλυση πρέπει να περιλαμβάνει το μέγεθος του τμήματος, την αναμενόμενη ανάπτυξη, τη συχνότητα αγοράς, την τρέχουσα χρήση του brand, την αφοσίωση στο brand και τις μακροπρόθεσμες δυνατότητες πωλήσεων και κερδών. Αυτές οι πληροφορίες μπορούν στη συνέχεια να χρησιμοποιηθούν για την κατάταξη των δυνητικών τμημάτων της αγοράς με βάση την ευκαιρία κέρδους, τον κίνδυνο, τη συνέπεια με τους στόχους της εταιρείας και άλλους παράγοντες που είναι σημαντικοί για την εταιρεία.

Το πέμπτο βήμα είναι η "επιλογή των αγορών-στόχων". Το βήμα αυτό δεν αποτελεί μέρος της διαδικασίας τμηματοποίησης αλλά φυσικό αποτέλεσμα αυτής. Είναι μια σημαντική απόφαση που επηρεάζει και συχνά καθορίζει άμεσα το μείγμα μάρκετινγκ της επιχείρησης. Το τελευταίο είναι "ο σχεδιασμός, η εφαρμογή και η διατήρηση του κατάλληλου μείγματος μάρκετινγκ". Το μείγμα μάρκετινγκ έχει περιγραφεί ως η στρατηγική που παίρνει η εταιρεία για το προϊόν, τη διανομή, τη προώθηση και την τιμή αυτού.

Οι ραγδαίες εξελίξεις στην τεχνολογία των υπολογιστών και η έκρηξη των συστημάτων συλλογής δεδομένων κατά τη διάρκεια των τελευταίων τριάντα χρόνων καθιστούν πιο κρίσιμη για τις επιχειρήσεις την κατανόηση των πελατών τους. Οι εταιρείες που χρησιμοποιούν δεδομένα βασίζονται σε στρατηγικές ανάλυσης, συχνά απολαμβάνουν ανταγωνιστικό πλεονέκτημα (Cross & Thompson, 2008).

Η ομαδοποίηση και η δημιουργία προφίλ της πελατειακής βάσης μπορούν να απαντήσουν τις ακόλουθες ερωτήσεις για την επιχείρηση (Cross & Thompson, 2008):

1. Ποιοι είναι οι πελάτες της;
2. Πόσο κερδοφόροι είναι οι πελάτες της;
3. Ποιοι είναι οι λιγότερο κερδοφόροι πελάτες της;
4. Γιατί φεύγουν οι πελάτες της;
5. Ποιο είναι το προφίλ των καλύτερων πελατών της;

Σύμφωνα με τους Kotler και Keller (2016), η τμηματοποίηση στο μάρκετινγκ μπορεί να χρησιμοποιηθεί για τα ακόλουθα:

- Μεγαλύτερη κατανόηση των πελατών για την υποστήριξη στον εντοπισμό νέων ευκαιριών μάρκετινγκ
- Σχεδιασμός και ανάπτυξη νέων προϊόντων/υπηρεσιών προσαρμοσμένα στα χαρακτηριστικά κάθε τμήματος και στην αγορά ως σύνολο
- Σχεδιασμός στρατηγικών προώθησης προϊόντων σε υπάρχον πελάτες σύμφωνα με τις διαπιστωμένες επιθυμίες και ανάγκες κάθε τμήματος
- Προσφορά προσαρμοσμένων ανταμοιβών και κινήτρων
- Επιλογή του κατάλληλου διαφημιστικού και επικοινωνιακού μηνύματος και καναλιού.
- Επιλογή του κατάλληλου καναλιού πωλήσεων και υπηρεσιών.
- Καθορισμός της εικόνας του brand και των βασικών πλεονεκτημάτων του προϊόντος που πρέπει να επικοινωνηθούν με βάση τα ιδιαίτερα χαρακτηριστικά κάθε τμήματος.
- Διαφοροποίηση στην εξυπηρέτηση πελατών ανάλογα με τις ιδιαιτερότητες του κάθε τμήματος.
- Αποτελεσματικότερη κατανομή των πόρων ανάλογα με τη δυνητική απόδοση του κάθε τμήματος.
- Ιεράρχηση των πρωτοβουλιών μάρκετινγκ που αποσκοπούν στη διατήρηση και ανάπτυξη πελατών ανάλογα με την αξία κάθε τμήματος.

2.4 Τύποι Τμηματοποίησης Πελατών

Υπάρχουν διάφορα κριτήρια για την τμηματοποίηση των πελατών που μπορούν να χρησιμοποιηθούν για τη βελτιστοποίηση των ενεργειών μάρκετινγκ προς τους καταναλωτές της εταιρείας. Παρουσιάζονται τρεις κύριοι τύποι τμηματοποίησης.

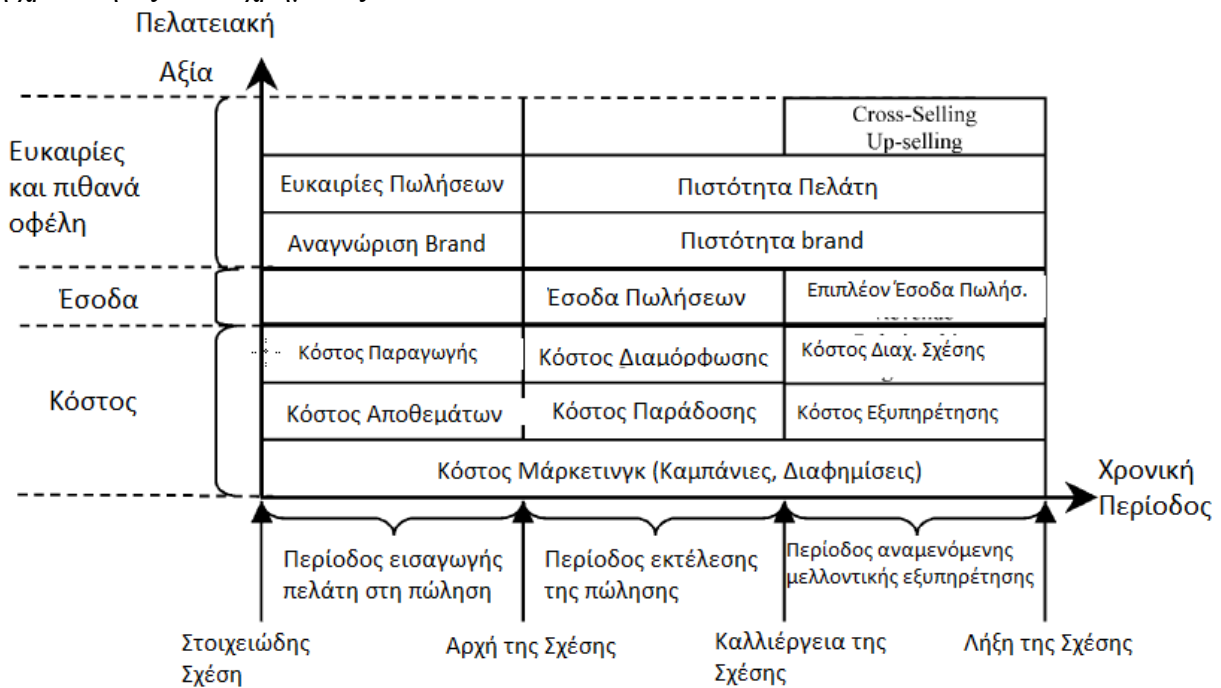
2.4.1 Με Βάση την Αξία (Value Based)

Στην τμηματοποίηση πελατών με βάση την αξία, οι πελάτες ομαδοποιούνται ανάλογα με την αξία τους. Πρόκειται για έναν από τους σημαντικότερους τύπους τμηματοποίησης, δεδομένου ότι μπορεί να χρησιμοποιηθεί για τον εντοπισμό των πιο πολύτιμων πελατών και για την παρακολούθηση της αξίας και των μεταβολών της αξίας με την πάροδο του χρόνου. Χρησιμοποιείται επίσης για τη διαφοροποίηση των στρατηγικών παροχής υπηρεσιών και για τη

βελτιστοποίηση της κατανομής των πόρων στις πρωτοβουλίες μάρκετινγκ. Η τμηματοποίηση με βάση την αξία είναι η διαδικασία διαχωρισμού της πελατειακής βάσης ανάλογα με την αξία που έχει ο κάθε πελάτης για την εταιρεία όπου βασίζεται σε διάφορους παράγοντες. Θα πρέπει να τονιστεί ότι αυτό δεν είναι μια εφάπαξ εργασία. Είναι ζωτικής σημασίας για τον οργανισμό να είναι σε θέση να παρακολουθεί τις μεταβολές της αξίας διαχρονικά. Ο οργανισμός θα πρέπει να παρακολουθεί και, αν είναι δυνατόν, να παρεμβαίνει προκειμένου να αποτρέψει τις μετακινήσεις της αξίας προς τα κάτω και να ενθαρρύνει τις μετακινήσεις της αξίας προς τα πάνω (Ziafat & Shakeri, 2014).

2.4.1.1 Life Time Value – Αξία Χρόνου Ζωής (LTV)

Ο Dwyer (1997) ορίζει το LTV ως το άθροισμα των εσόδων που αποκτώνται από τους πελάτες της εταιρείας κατά τη διάρκεια της ζωής των συναλλαγών των πελατών, μετά την αφαίρεση του συνολικού κόστους προσέλκυσης, πώλησης και εξυπηρέτησης των πελατών, λαμβάνοντας υπόψη τη χρονική αξία του χρήματος.



Σχήμα 2: Δομή LTV προς τη πάροδο του χρόνου (Hwang, Jung, & Suh, 2004)

Ο οριζόντιος άξονας του σχήματος 2 υποδηλώνει το είδος της σχέσης ανάλογα με το χρονικό πλαίσιο, ενώ ο κάθετος, το είδος της αξίας του πελάτη προς μια εταιρεία. Μια εταιρεία διαμορφώνει διάφορες σχέσεις ανάλογα με τα στάδια της σχέσης - στοιχειώδες, αρχικό, προωθητικό και στάδιο λήξης. Ένας πελάτης δίνει επίσης σε μια εταιρεία διάφορα έσοδα, κόστη και ευκαιρίες, αλλά και πιθανά οφέλη (Hwang et al. 2004).

2.4.1.2 Customer Lifetime Value – Αξία Χρόνου Ζωής Πελάτη (CLV)

Η κίνηση προς μια πελατοκεντρική προσέγγιση του μάρκετινγκ, σε συνδυασμό με την αυξανόμενη διαθεσιμότητα δεδομένων συναλλαγών πελατών, έχει οδηγήσει σε ένα ενδιαφέρον τόσο για την έννοια όσο και για τον υπολογισμό της αξίας χρόνου ζωής των πελατών (CLV). Σε καθαρά εννοιολογικό επίπεδο, ο υπολογισμός της CLV είναι μια απλή πρόταση: Είναι απλώς η παρούσα αξία των μελλοντικών ταμειακών ροών που συνδέονται με έναν πελάτη (Fader et al. 2005).

Ως κερδοφόρος πελάτης ορίζεται ένα άτομο, νοικοκυριό ή εταιρεία, τα έσοδα του οποίου σε βάθος χρόνου υπερβαίνουν, κατά ένα αποδεκτό ποσό, το κόστος προσέλκυσης, πώλησης και εξυπηρέτησης του εν λόγω πελάτη. Το πλεόνασμα ονομάζεται CLV. Τα μοντέλα CLV μπορούν να είναι χρήσιμα για τον προσδιορισμό των χαρακτηριστικών των πελατών μιας επιχείρησης, ενώ η τμηματοποίηση των πελατών με βάση αυτά τα χαρακτηριστικά/μεταβλητές, μπορεί να βοηθήσει στην κατανομή των πόρων για διάφορες δραστηριότητες μάρκετινγκ (Kotler et al. 2005).

Τα διάφορα οφέλη της CLV είναι τα εξής:

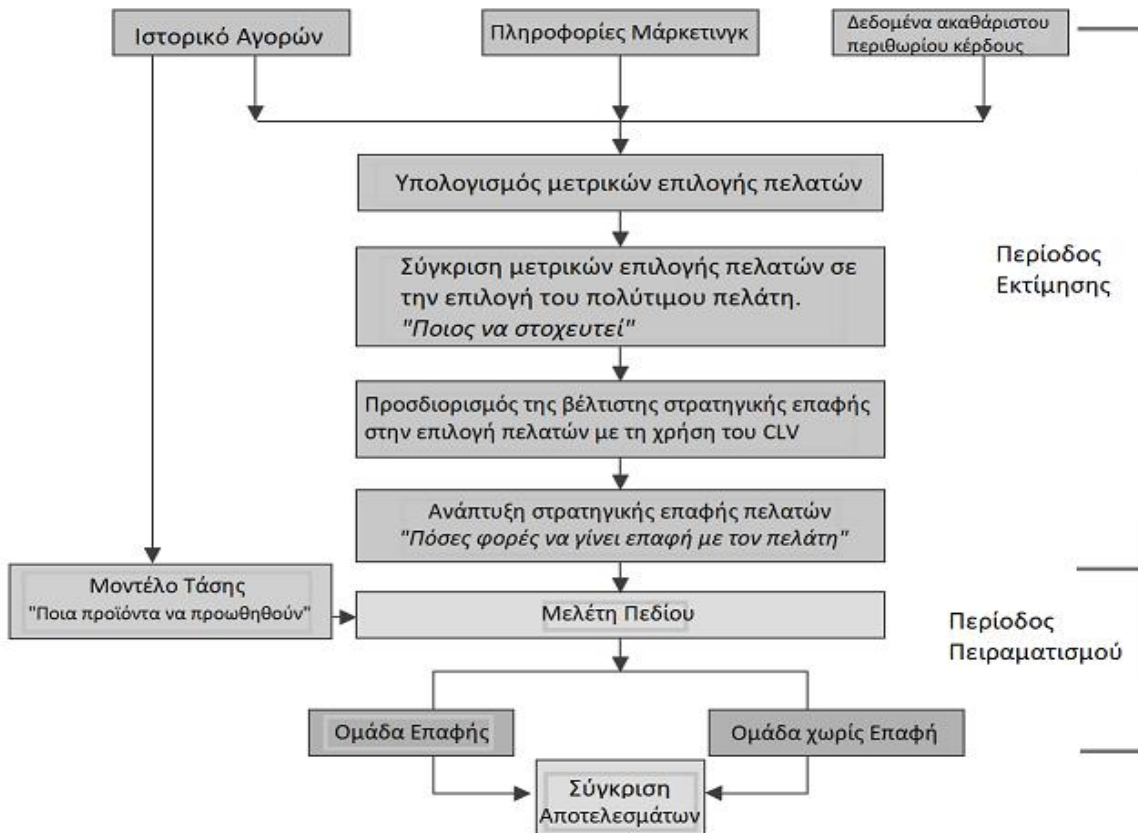
- Η εστίαση στην CLV βοηθά στη λήψη αποφάσεων σχετικά με τα δεδομένα που πρέπει να συλλεχθούν σε επιχειρησιακό επίπεδο, καθώς δεν υπάρχει έλλειψη διαθέσιμων πληροφοριών σήμερα. Οι πληροφορίες αυτές θα πρέπει να μετατραπούν σε πληροφορίες για την υποστήριξη στη λήψη αποφάσεων.
- Αναδεικνύει την αξία των διαφόρων περιουσιακών στοιχείων για την επιχείρηση. Τα μη σχετιζόμενα περιουσιακά στοιχεία είναι πολύτιμα στο βαθμό που αυξάνουν την αξία των περιουσιακών στοιχείων των πελατών
- Ο ρόλος του τμήματος μάρκετινγκ επαναπροσδιορίζεται έτσι ώστε να διαχειρίζεται τα στοιχεία των πελατών με τρόπο που να διευκολύνει την πελατειακή αξία
- Σε οργανωτικό επίπεδο, παρέχονται μετρήσεις για την εκτίμηση της επίδρασης των διαφόρων οργανωτικών δραστηριοτήτων στην αξία των πελατών.
- Βοηθά στην ανάπτυξη μιας πιο βασισμένης στην αξία τμηματοποίησης των πελατών (Hogan et al. 2002).

Η απόκτηση πελατών, η διατήρηση και οι διασταυρούμενες πωλήσεις καθορίζουν τη μακροχρόνια κερδοφορία ή την αξία χρόνου ζωής ενός πελάτη. Οι (Gupta & Zeithaml, 2006) ορίζουν την CLV ως τη παρούσα αξία όλων των μελλοντικών κερδών που αποκτώνται από έναν πελάτη κατά τη διάρκεια της ζωής της σχέσης του με μια επιχείρηση. Η CLV είναι παρόμοια με την προσέγγιση των προεξοφλημένων ταμειακών ροών που χρησιμοποιείται στα χρηματοοικονομικά, εκτός από δύο βασικές διαφορές. Πρώτον, η CLV ορίζεται σε επίπεδο μεμονωμένου πελάτη ή τμήματος, αναγνωρίζοντας ότι ορισμένοι πελάτες είναι πιο σημαντικοί και κερδοφόροι από άλλους. Δεύτερον, η CLV ενσωματώνει ρητά την πιθανότητα ένας πελάτης να αποστατήσει από τους ανταγωνιστές στο μέλλον. Η CLV είναι η μακροπρόθεσμη κερδοφορία ενός μεμονωμένου πελάτη. Είναι χρήσιμη για τον προσδιορισμό των πελατών που πρέπει να αποκτηθούν, του ποσού που πρέπει να δαπανηθεί για αυτούς και του τρόπου προσαρμογής του μάρκετινγκ και των προσφορών προϊόντων σε αυτούς.

Ο Kumar et al. (2008) προτείνουν το εξής πλαίσιο του πίνακα 1 για τη διαχείριση του CLV, και στο σχήμα 3 απεικονίζεται η εφαρμογή αυτού:

Πίνακας 1: Πλαίσιο Διαχείρισης CLV (Kumar, Venkatesan, Bohling, & Beckmann, 2008)

Διαδικασία	Σκοπός
Μέτρηση Αξίας Χρόνου Ζωής Πελάτη	Για να ληφθεί ένα μέτρο της δυνητικής αξίας των πελατών της εταιρείας
Προσδιορισμός των παραγόντων του CLV	Έτσι ώστε οι μάνατζερ να μπορούν να επηρεάσουν το CLV
Καθορισμός του βέλτιστου επιπέδου επαφών για κάθε πελάτη που θα μεγιστοποιούσε το αντίστοιχο CLV του.	Να καθοδηγήσει τους μάνατζερ σχετικά με το επίπεδο της επένδυσης που απαιτείται για κάθε πελάτη
Ανάπτυξη μοντέλων τάσης για την πρόβλεψη των προϊόντων που είναι πιθανό να αγοράσει ένας πελάτης	Για την ανάπτυξη ενός μηνύματος προϊόντος κατά την επικοινωνία με τον πελάτη
Ανακατανομή των επαφών μάρκετινγκ από πελάτες χαμηλού CLV σε πελάτες υψηλού CLV.	Μεγιστοποίηση της παραγωγικότητας του μάρκετινγκ



Σχήμα 3: Εφαρμογή του πλαισίου διαχείρισης της αξίας χρόνου ζωής του πελάτη (Kumar, Venkatesan, Bohling, & Beckmann, 2008)

2.4.2 Με βάση τη συμπεριφορά (Behavioral Based)

Σε αυτή την τμηματοποίηση, οι καταναλωτές χωρίζονται σε ομάδες με βάση τη γνώση, τη στάση, τη χρήση ή την ανταπόκρισή τους σε ένα προϊόν. Ο Kotler (2003) διατυπώνει πως "Πολλά στελέχη μάρκετινγκ πιστεύουν ότι οι μεταβλητές συμπεριφοράς, συμπεριλαμβανομένων των περιστάσεων, των πλεονεκτημάτων, της κατάστασης χρήστη, του ποσοστού χρήσης και της κατάστασης αφοσίωσης, είναι τα καλύτερα σημεία εκκίνησης για την κατασκευή τμημάτων της αγοράς".

Οι καταναλωτές μπορούν αρχικά να διακριθούν ανάλογα με τις περιστάσεις κατά τις οποίες αρχίζουν να χρειάζονται, να αγοράζουν ή να χρησιμοποιούν τα προϊόντα της επιχειρήσεως. Η τμηματοποίηση κατά περίπτωση μπορεί να βοηθήσει τις εταιρείες να διευρύνουν το πεδίο χρήσης των προϊόντων τους (Meredith et al. 2002).

Μια άλλη διάσταση της συμπεριφορικής τμηματοποίησης είναι η κατάσταση του χρήστη. Οι αγορές μπορούν να τμηματοποιηθούν σε ομάδες μη χρηστών, πρώην χρηστών, δυνητικών χρηστών, χρηστών που χρησιμοποιούν για πρώτη φορά το προϊόν/την υπηρεσία και τακτικών χρηστών ενός προϊόντος ή μιας υπηρεσίας (Kotler et al. 2005).

Το τελευταίο είναι το καθεστώς αφοσίωσης. Οι καταναλωτές μπορεί να είναι πιστοί σε ορισμένες μάρκες, καταστήματα ή εταιρείες. Ο Rossi et al. (1996) ορίζουν την πελατειακή αφοσίωση ως μια βαθιά υπόσχεση που δεν μπορεί να επηρεαστεί από το εξωτερικό περιβάλλον ή από τις μετασχηματισμένες δραστηριότητες προώθησης και θα συνεχίσει να προτιμά κάποιο προϊόν ή υπηρεσία και να αγοράζει ξανά και ξανά. Οι αγοραστές μπορούν να χωριστούν σε τέσσερις ομάδες ανάλογα με την κατάσταση αφοσίωσης στη μάρκα: σκληροπυρηνικής πιστότητας (καταναλωτές που αγοράζουν πάντα μια μάρκα), διαιρεμένης πιστότητας (καταναλωτές που είναι πιστοί σε δύο ή τρεις μάρκες), μεταβαλλόμενης πιστότητας (καταναλωτές που μετακινούνται από μια μάρκα σε μια άλλη), και εναλλασσόμενοι (καταναλωτές που δεν δείχνουν καμία αφοσίωση σε καμία μάρκα).

Οι επιχειρήσεις μπορούν να μάθουν πολλά από την ανάλυση του βαθμού αφοσίωσης των καταναλωτών. Πρώτον, οι επιχειρήσεις μπορούν να γνωρίζουν ποιοι καταναλωτές είναι πιστοί στα προϊόντα τους. Εάν μελετήσουν εκείνους τους καταναλωτές με χαμηλό βαθμό αφοσίωσης, μπορούν να μάθουν ποιες μάρκες είναι οι πραγματικοί ανταγωνιστές τους. Δεύτερον, μέσω της μελέτης εκείνων των καταναλωτών που έχουν εγκαταλείψει τα προϊόντα τους, οι εταιρείες μπορούν να γνωρίσουν τα αδύνατα σημεία της στρατηγικής μάρκετινγκ. Και επιπλέον, όσον αφορά εκείνους τους καταναλωτές χωρίς αφοσίωση, οι επιχειρήσεις μπορούν να κάνουν χρήση της προώθησης για να τους προσελκύσουν (Sun, 2009).

2.4.3 Με βάση τα δημογραφικά (Demographics Based)

Η δημογραφική τμηματοποίηση φαίνεται να είναι η πιο διαδεδομένη μορφή τμηματοποίησης της αγοράς. Αυτό οφείλεται πιθανώς στο γεγονός ότι οι καταναλωτές τοποθετούνται σε συγκεκριμένες κλίμακες μέτρησης που είναι εύκολα κατανοητές. Οι πληροφορίες ερμηνεύονται εύκολα, συλλέγονται σχετικά εύκολα και μεταφέρονται εύκολα από τη μία μελέτη στην άλλη. Οι συνήθεις δημογραφικές μεταβλητές είναι η ηλικία, το φύλο, το μέγεθος και ο τύπος της οικογένειας, το εισόδημα, το μορφωτικό επίπεδο, η

φυλή και η εθνικότητα. Μερικές φορές χρησιμοποιούνται συνδυασμοί αυτών των μεταβλητών ανάλογα με το βαθμό εξειδίκευσης που απαιτείται για την κατασκευή του τμήματος (Beane & Ennis, 1987).

Οι ανάγκες και οι επιθυμίες των καταναλωτών αλλάζουν με την ηλικία. Ορισμένες εταιρείες χρησιμοποιούν την τμηματοποίηση κατά ηλικία και κύκλου ζωής, προσφέροντας διαφορετικά προϊόντα ή χρησιμοποιώντας διαφορετικές προσεγγίσεις μάρκετινγκ για διαφορετικές ομάδες ηλικίας και κύκλου ζωής (Kotler et al. 2005). Σε ορισμένες περιπτώσεις η έννοια του κύκλου ζωής αποδείχθηκε πιο χρήσιμη μεταβλητή τμηματοποίησης από την ηλικία. Η βασική παραδοχή που διέπει την προσέγγιση του οικογενειακού κύκλου ζωής είναι ότι τα περισσότερα νοικοκυριά περνούν από μια ομαλή εξέλιξη σταδίων το καθένα με τα δικά του χαρακτηριστικά αγοραστικά πρότυπα. Παρά τις δυσκολίες ταξινόμησης ορισμένων καταναλωτών που δεν εντάσσονται με σαφήνεια σε κανένα από τα συνήθη στάδια, π.χ. οι ηλικιωμένοι άγαμοι ή η χήρα που έχει μικρά παιδιά, ο κύκλος ζωής παραμένει μια χρήσιμη βάση τμηματοποίησης σε γενικό επίπεδο (Tynan & Drayton, 1987).

Σε πολλούς τομείς προϊόντων το φύλο του καταναλωτή καθορίζει το προϊόν που θα αγοράσει. Για παράδειγμα, η τμηματοποίηση κατά φύλο εφαρμόζεται εδώ και πολύ καιρό στα είδη ένδυσης, στα περιοδικά, στα καλλυντικά και στα είδη υγιεινής (Tynan & Drayton, 1987).

Η τμηματοποίηση του εισοδήματος χρησιμοποιείται συχνά για τη προώθηση προϊόντων και υπηρεσιών όπως αυτοκίνητα, σκάφη, ρούχα, καλλυντικά και ταξίδια. Πολλές εταιρείες απευθύνονται σε εύπορους καταναλωτές για τη προώθηση πολυτελών αγαθών και υπηρεσιών ευκολίας (Kotler et al. 2005).

2.5 RFM Ανάλυση (Recency, Frequency, Monetary value)

Η ανάλυση RFM προέρχεται από την πρακτική του άμεσου μάρκετινγκ σε εταιρείες πωλήσεων καταλόγων τη δεκαετία του 1960. Η ανάλυση αυτή προβλέπει μια τμηματοποίηση των πελατών στη βάση δεδομένων της εταιρείας με βάση τη συμπεριφορά του παρελθόντος (Hughes, 2000) & (Coussement et al. 2014).

Η ανάλυση RFM, ορίζεται ως Recency (R) “πόσο πρόσφατα έγινε αγορά από τον πελάτη”, Frequency (F) “πόσο συχνά γίνονται αγορές από τον πελάτη” και Monetary value (M) “ποια η χρηματική αξία της αγοράς του πελάτη” και προσδιορίζει τους πελάτες που είναι πιο πιθανό να αγοράσουν ξανά επειδή έχουν αγοράσει πρόσφατα, αγοράζουν συχνά ή έχουν ξοδέψει ένα συγκεκριμένο χρηματικό ποσό στην επιχείρηση (Lamb et al. 2011).

Σύμφωνα με τους Tsiptsis & Chorianoopoulos (2009), οι διαστάσεις της ανάλυσης RFM στο λιανικό εμπόριο, συχνά είναι οι εξής:

- **Recency (Επικαιρότητα):** Ο χρόνος (σε μονάδες όπως ημέρες/μήνες/έτη) από την πιο πρόσφατη συναλλαγή αγοράς ή επίσκεψη στο κατάστημα για αγορά
- **Frequency (Συχνότητα):** Ο συνολικός αριθμός των συναλλαγών αγοράς ή των επισκέψεων σε καταστήματα κατά την εξεταζόμενη περίοδο. Μια εναλλακτική, και πιθανώς καλύτερα καθορισμένη, προσέγγιση που λαμβάνει επίσης υπόψη τη διάρκεια ζωής του πελάτη και υπολογίζει τη συχνότητα ως το μέσο αριθμό συναλλαγών ανά μονάδα χρόνου, για παράδειγμα το μηνιαίο μέσο αριθμό συναλλαγών
- **Monetary value (Χρηματική αξία):** Η συνολική αξία των αγορών εντός της εξεταζόμενης περιόδου ή η μέση αξία (π.χ. μέση μηνιαία αξία) ανά χρονική μονάδα. Σύμφωνα με έναν εναλλακτικό, αλλά όχι τόσο δημοφιλή, ορισμό, ο νομισματικός δείκτης ορίζεται ως η μέση αξία συναλλαγής (μέση αξία ανά συναλλαγή αγοράς). Δεδομένου ότι η συνολική αξία τείνει να συσχετίζεται με τη συχνότητα των συναλλαγών, το σκεπτικό πίσω από αυτόν τον εναλλακτικό ορισμό είναι να αποτυπωθεί μια διαφορετική και συμπληρωματική πτυχή της αγοραστικής συμπεριφοράς.

Οι πελάτες που αγοράζουν συχνά, ξοδεύουν σημαντικές ποσότητες χρήματων και η τελευταία τους αγορά ήταν πρόσφατη, είναι πιο πιθανό να αγοράσουν ξανά. Οι επιχειρήσεις αναπτύσσουν εξισώσεις για να εντοπίσουν τους "καλύτερους πελάτες" (συνήθως το 20 % της πελατειακής βάσης) αποδίδοντας μια βαθμολογία στις εγγραφές πελατών στη βάση δεδομένων σχετικά με το πόσο συχνά, πόσο πρόσφατα και πόσο έχουν ξοδέψει. Στη συνέχεια, οι πελάτες κατατάσσονται για να καθοριστεί ποιοι από αυτούς μετακινούνται στην κορυφή του καταλόγου και ποιοι κατεβαίνουν στη κατάταξη. Η κατάταξη παρέχει τη βάση για τη μεγιστοποίηση των κερδών, διότι επιτρέπει στην επιχείρηση να χρησιμοποιήσει τις πληροφορίες της βάσης δεδομένων των πελατών της για να επιλέξει εκείνα τα άτομα που έχουν αποδειχθεί καλές πηγές εσόδων (Lamb et al. 2011).

2.5.1 Αξία ανάλυσης RFM

Η ανάλυση RFM υπάρχει και χρησιμοποιείται στις επιχειρήσεις για περισσότερα από πενήντα χρόνια. Οι Hughes και Sellers (2004) αναφέρουν πως η αξία της ανάλυσης RFM είναι γνωστή και εφαρμόζεται ευρέως σήμερα, ως μέθοδος εντοπισμού των πελατών που έχουν υψηλή ανταπόκριση σε προωθητικές ενέργειες μάρκετινγκ και για τη βελτίωση των συνολικών ποσοστών ανταπόκρισης. Οι πελάτες που έχουν αγοράσει από την επιχείρηση πρόσφατα είναι πιο πιθανό να ανταποκριθούν στην επόμενη προώθησή της από εκείνους των οποίων η τελευταία αγορά ήταν πιο μακριά στο παρελθόν. Αυτή είναι μια καθολική αρχή που έχει διαπιστωθεί ότι ισχύει σε όλους σχεδόν τους κλάδους: ασφάλειες, τράπεζες, λιανικό εμπόριο, ταξίδια κ.λπ. Ισχύει επίσης ότι οι συχνοί αγοραστές είναι πιο πιθανό να ανταποκριθούν από τους λιγότερο συχνούς αγοραστές. Τέλος, οι πελάτες που ξοδεύουν μεγάλα ποσά συχνά ανταποκρίνονται καλύτερα από αυτούς που ξοδεύουν λιγότερα. Αυτές είναι οι τρεις απλές αρχές που κρύβονται πίσω από την ανάλυση RFM.

Οι επιμέρους συνιστώσες της RFM και οι παραγόμενες τμηματοποιήσεις, παρέχουν χρήσιμες πληροφορίες σχετικά με τις αγοραστικές συνήθειες των καταναλωτών. Αναμφίβολα, κάθε επιχείρηση λιανικής πώλησης θα πρέπει να παρακολουθεί τη συχνότητα, την ένταση και την επανάληψη των αγορών, καθώς αποτελούν σημαντικές διαστάσεις της σχέσης του πελάτη με την επιχείρηση. Επιπλέον, παρακολουθώντας τις μεταβολές της RFM με την πάροδο του χρόνου, ένας οργανισμός μπορεί να παρακολουθεί τις αλλαγές στις αγοραστικές συνήθειες κάθε πελάτη και να χρησιμοποιεί αυτές τις πληροφορίες για να ενεργοποιεί προληπτικά τις κατάλληλες ενέργειες μάρκετινγκ. Για παράδειγμα, συγκεκριμένα γεγονότα, όπως η μείωση της συνολικής αξίας των αγορών, η ξαφνική μείωση της συχνότητας των επισκέψεων ή η απουσία εμφανίσεων για ασυνήθιστα μεγάλο χρονικό διάστημα, μπορεί να υποδηλώνουν την αρχή του τέλους της σχέσης του πελάτη με τον οργανισμό. Αυτά τα σήματα, εάν αναγνωριστούν εγκαίρως, θα πρέπει να δρομολογήσουν εκστρατείες για την διατήρηση και την επανενεργοποίηση του πελάτη (Tsiptsis & Chorianoopoulos, 2009).

Ο στόχος της ανάλυσης RFM είναι να προσδιοριστεί ένα τμήμα πελατών που έχουν μεγάλη πιθανότητα να ανταποκριθούν σε μια εκστρατεία μάρκετινγκ. Εστιάζοντας σε αυτούς τους πελάτες και αποφεύγοντας να δαπανήσει πόρους σε πελάτες που δεν θα είχαν ανταποκριθεί ούτε ή άλλως, μια εταιρεία κάνει τις ενέργειες μάρκετινγκ της πιο στοχευμένη (Coussement et al. 2014).

Εκτός από τα πολύτιμα πλεονεκτήματα της ανάλυσης RFM, υπάρχουν αρκετά σημαντικά μειονεκτήματα χωρίζοντας το δείγμα πελατών σε ίσα μέρη των 20%.

Πρώτον, η διαδικασία διακριτοποίησης εισάγει μια απώλεια επεξηγηματικών πληροφοριών. Δεύτερον, η διαδικασία κωδικοποίησης πελατών είναι αυθαίρετη. Ανάλογα με την περίπτωση, περισσότερες ή λιγότερες κατηγορίες μπορεί να είναι καταλληλότερες. Για παράδειγμα, ανάλογα με τον προϋπολογισμό, θα μπορούσαν να χρησιμοποιηθούν λεπτότερα ή πιο χονδροειδή σχήματα κωδικοποίησης RFM. Επίσης, η τεχνική δεν είναι κατάλληλη για την προσθήκη άλλων

χαρακτηριστικών που μπορεί να σχετίζονται με τη μελλοντική συμπεριφορά ανταπόκρισης ενός πελάτη (Hughes, 2000).

Τέλος, Ένα προφανές μειονέκτημα αυτής της προσέγγισης είναι ότι συνήθως καταλήγει με σχεδόν την ίδια λίστα στόχου καλών πελατών, οι οποίοι θα μπορούσαν να ενοχληθούν από τις επαναλαμβανόμενες επαφές AM (Άμεσου Μάρκετινγκ). Αν και χρήσιμη, η προσέγγιση RFM, όταν δεν συνδυάζεται με άλλα σημαντικά χαρακτηριστικά των πελατών, όπως οι προτιμήσεις των προϊόντων, αποτυγχάνει να παράσχει πλήρη κατανόηση της συμπεριφοράς των πελατών. Μια επιχείρηση θα πρέπει να έχει μια πλήρη εικόνα του πελάτη και να χρησιμοποιεί όλες τις διαθέσιμες πληροφορίες για να κατευθύνει την επιχειρηματικές αποφάσεις (Tsiptsis & Chorianopoulos, 2009).

Όπως αναφέρει ο Hughes (2005), η ανάλυση RFM δεν πρέπει να χρησιμοποιείται πάντα. Αν ναι, οι μισοί πελάτες της επιχείρησης δεν θα λαμβάνουν ενέργειες AM, και θα πάνε να είναι πελάτες. Πρέπει να χρησιμοποιείται σε αραιά χρονικά διαστήματα για την καλύτερη κατανόηση της πελατειακής βάσης και όταν η επιχείρηση χρειάζεται μια αύξηση στα κέρδη της.

2.5.2 Χρήση ανάλυσης RFM

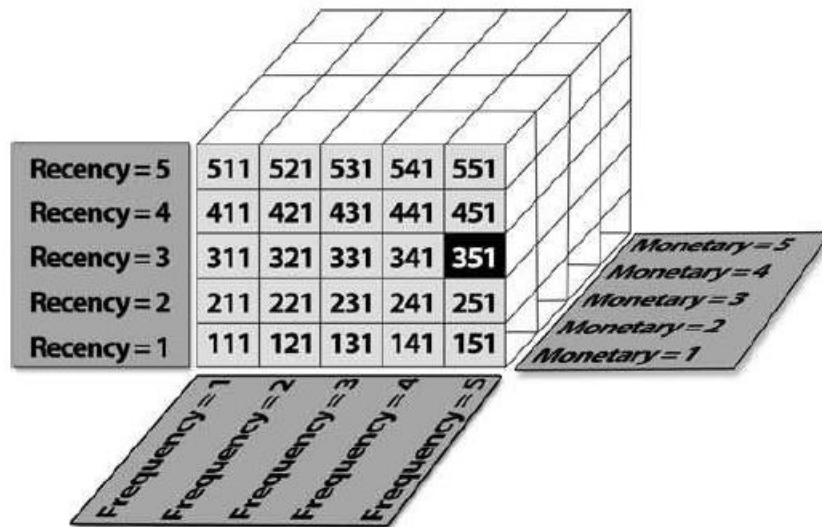
Τα περισσότερα στελέχη άμεσου μάρκετινγκ (Direct Marketing) εφαρμόζουν τον τύπο RFM για να επιλέξουν πελάτες ανάλογα με το πόσος χρόνος έχει περάσει από την τελευταία τους αγορά, πόσες φορές έχουν αγοράσει και πόσα έχουν ξοδέψει από τότε που έγιναν πελάτες. Έστω ότι μια εταιρεία θα προωθήσει την προσφορά ένα δερμάτινου μπουφάν. Θα μπορούσε να κάνει αυτή την προσφορά στους πιο πολύτιμους πελάτες - αυτούς που έκαναν την τελευταία τους αγορά πριν από 30 έως 60 ημέρες, που πραγματοποιούν τρεις έως έξι αγορές το χρόνο και που έχουν ξοδέψει τουλάχιστον 100 δολάρια από τότε που έγιναν πελάτες. Καθορίζονται πόντοι στους πελάτες για τα διαφορετικά επίπεδα RFM - όσο περισσότερους πόντους συλλέγει ο πελάτης, τόσο πιο πολύτιμος (Kotler & Keller, 2016).

Τα έμπειρα στελέχη άμεσου μάρκετινγκ παίρνουν τις 3 αρχές της ανάλυσης RFM και τις ποσοτικοποιούν. Κωδικοποιούν όλους τους πελάτες σε κελιά RFM και εξετάζουν τα ποσοστά ανταπόκρισης των πελατών σε κάθε κελί όταν εκτίθενται στην ίδια προώθηση. Είναι αλήθεια, βέβαια, ότι μόνο ένα ποσοστό των πελατών θα πραγματοποιήσει μια πρόσθετη αγορά με βάση μια νέα προώθηση. Αλλά, από εκείνους που ανταποκρίνονται, οι απαντήσεις προέρχονται συνήθως από πελάτες σε κελιά RFM υψηλότερης κατάταξης (Hughes & Sellers, 2004).

2.5.3 Μοντέλο ανάλυσης RFM

Το μοντέλο που προτάθηκε από τον Hughes (2000), πρωτοπόρο στον τομέα του μάρκετινγκ βάσεων δεδομένων, τους Tsiptsis και Chorianopoulos (2009), αλλά και τους Blattberg et al. (2008)

ορίζουν πως οι μεταβλητές RFM μετατρέπονται σε διακριτές τιμές {1, 2, 3, 4, 5}. Έτσι, κάθε πελάτης λαμβάνει μια τιμή για κάθε μεταβλητή RFM. Αναλυτικότερα, οι αποδιδόμενες τιμές προέρχονται από την ακόλουθη διαδικασία. Το πρώτο βήμα συνοψίζει το ιστορικό αγορών των πελατών σε μεταβλητές RFM. Αυτές οι πληροφορίες αγοραστικού ιστορικού μπορούν εύκολα να προκύψουν από τη βάση δεδομένων συναλλαγών μιας εταιρείας, στην οποία καταγράφονται όλες οι αγορές των πελατών. Το δεύτερο βήμα ταξινομεί τους πελάτες με βάση την επικαιρότητα (πόσο πρόσφατη ήταν η τελευταία τους αγορά), και τους χωρίζει σε 5 ίσες ομάδες πελατών και τους αποδίδει μία από τις πέντε διακριτές τιμές. Για παράδειγμα, το 20% των πελατών που αγόρασαν πιο πρόσφατα λαμβάνουν τον κωδικό 5. Το τρίτο βήμα, εντός κάθε ομάδας επικαιρότητας, ταξινομεί τους πελάτες με βάση τη συχνότητα αγορών τους, αποδίδοντας τους τιμές με παρόμοιο τρόπο. Τέλος, το τέταρτο βήμα ταξινομεί κάθε ομάδα συχνότητας με βάση τη χρηματική αξία των αγορών τους και αποδίδει και πάλι τιμές στα υποσύνολα. Από αυτή τη διαδικασία, κάθε πελάτης λαμβάνει τρεις τιμές που υποδηλώνουν την ένταξη σε μία από τις 125 (5×5×5) ομάδες ίσου μεγέθους, όπως φαίνεται και από το σχήμα 4 των Tsiptsis και Chorianopoulos (2009). Αυτή η κατάταξη επιτρέπει στους υπεύθυνους λήψης αποφάσεων μάρκετινγκ να διαμορφώσουν ένα σύνολο κανόνων που βοηθούν στον εντοπισμό των πελατών τους οποίους θα πρέπει να στοχεύσει μια μελλοντική εκστρατεία άμεσου μάρκετινγκ(AM).



Σχήμα 4: Το σύνολο των κελιών RFM στην περίπτωση διαχωρισμού σε πεντημόρια (ομάδες του 20%) (Tsiptsis & Chorianopoulos, 2009)

Για παράδειγμα, στην υπόθεση όπου ένα στέλεχος άμεσου μάρκετινγκ επιλέγει τυχαία 10.000 πελάτες από τη λίστα που έλαβαν πρόσφατα ένα μήνυμα AM. Τα αποτελέσματα στον πίνακα 2 αποτελούν ουσιαστικά μια διασταύρωση των στοιχείων R επί F επί M επί των ανταποκρίσεων των πελατών στην προωθητική καμπάνια AM, και αποτελούν το μοντέλο RFM για την πρόβλεψη του ποσοστού ανταπόκρισης των πελατών για μελλοντική καμπάνια. Για παράδειγμα, εάν ένας πελάτης έχει 5 και στις 3 τιμές R, F και M, η πιθανότητα απόκρισης προβλέπεται να είναι 15%. Αφού υπολογιστούν οι προβλεπόμενες πιθανότητες ανταπόκρισης για όλους τους πελάτες της λίστας, το στέλεχος μάρκετινγκ να επιλέξει μια ομάδα πελατών για μια επερχόμενη προωθητική καμπάνια AM (Blattberg et al. 2008).

Πίνακας 2: Οι προβλεπόμενες πιθανότητες απόκρισης από το μοντέλο RFM (Blattberg, Kim, & Neslin, 2008)

Διακριτές R F M	Τιμές	Πλήθος πελατών που δέχθηκαν μήνυμα AM	Πλήθος που ανταποκρίθηκε	Ποσοστό ανταπόκρισης
5 5 5		100	15	15.0%
5 5 4		90	13	14.4%
5 5 3		100	13	13.0%
5 5 2		80	10	12.5%
5 5 1		70	7	10.0%
4 5 1		100	8	8.0%
---		---	---	---
---		---	---	---
1 1 2		70	1	0.01%
1 1 1		90	0	0.00%
Σύνολο		10.000	350	3.50%

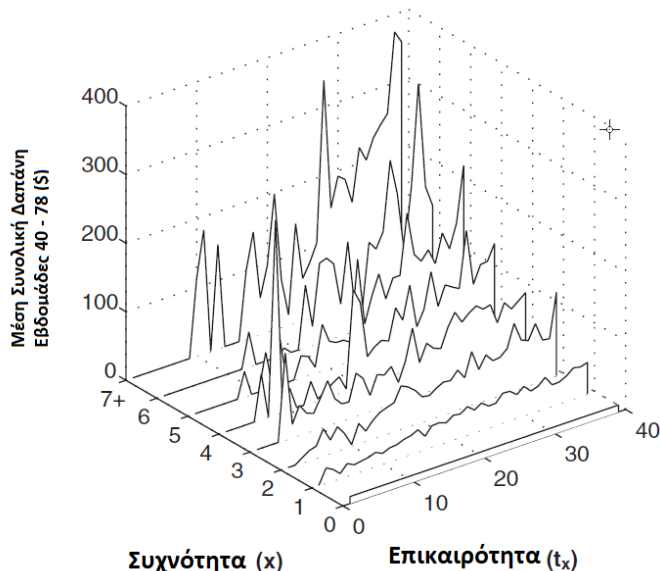
Αυτός ο συνδυασμός των στοιχείων R, F και M σε κύτταρα χρησιμοποιείται ευρέως, αν και έχει ένα συγκεκριμένο μειονέκτημα. Ο μεγάλος αριθμός των παραγόμενων κυττάρων καθιστά τη διαδικασία αρκετά δυσκίνητη και δύσκολα διαχειρίσιμη. Μια εναλλακτική μέθοδος για την τμηματοποίηση των πελατών σύμφωνα με τα μοτίβα RFM τους είναι να χρησιμοποιηθούν οι αντίστοιχες συνιστώσες ως είσοδοι σε ένα μοντέλο ομαδοποίησης και ο αλγόριθμος να αποκαλύψει τις υποκείμενες φυσικές ομαδοποιήσεις των πελατών (Tsiptsis & Chorianopoulos, 2009).

Στην πραγματική ζωή, διάφορα κριτήρια μπορούν να καθοδηγήσουν την επιλογή του σωστού αριθμού πελατών που πρέπει να στοχευθούν. Για παράδειγμα, ο αναλυτής θα μπορούσε να λάβει ένα αυθαίρετο ποσοστό του αρχείου πελατών ή να καθορίσει έναν βέλτιστο αριθμό πελατών-στόχων για τη μεγιστοποίηση των κερδών της εταιρείας. Επίσης, το μοντέλο ταξινομεί σε υψηλά δεκατημόρια τους πελάτες με ευνοϊκές βαθμολογίες RFM και, ως εκ τούτου, επιλέγονται για

επικοινωνία. Ως αποτέλεσμα, οι βαθμολογίες RFM τους βελτιώνονται, ενώ όσοι δεν έρχονται σε επαφή αναπτύσσουν χαμηλότερες βαθμολογίες RFM. Την επόμενη φορά που εφαρμόζεται το μοντέλο, οι ίδιοι πελάτες βαθμολογούνται ξανά στα υψηλά δεκατημόρια και οι ίδιοι πελάτες στα χαμηλά δεκατημόρια και ο κύκλος συνεχίζεται. Με την πάροδο του χρόνου, μια ομάδα πελατών λαμβάνει μεγάλη προσοχή- μια άλλη ομάδα λαμβάνει ελάχιστη. Η πρώτη ομάδα απαρτίζεται από τους καλύτερους πελάτες, αλλά η δεύτερη ομάδα χειροτερεύει με τη πάροδο του χρόνου (Blattberg et al. 2008).

2.5.4 Παραδείγματα χρήσης της RFM στη πραγματική ζωή

Οι Fader et al. (2005) χρησιμοποιώντας την ανάλυση RFM σε αγοραστικά δεδομένα της CDNOW, εταιρείας πώλησης δίσκων μουσικής, εντόπισαν πως υπάρχει μια εξαιρετικά μη γραμμική σχέση μεταξύ της επικαιρότητας / συχνότητας και των μελλοντικών συναλλαγών. Ειδικότερα, ανέλυσαν αγοραστικά δεδομένα 23,750 πελατών της CDNOW, από το πρώτο τρίμηνο του 1997 μέχρι και τον Ιούνιο του 1998 (78 εβδομάδες). Χωρίζοντας την περίοδο σε 2 ίσα κομμάτια (39 εβδομάδες), εφάρμοσαν το μοντέλο RFM στις πρώτες 39 εβδομάδες για να αναλύσουν αν υπάρχει συσχέτιση μεταξύ της συχνότητας και της επικαιρότητας απέναντι στις αγορές που έγιναν στις εβδομάδες 40 με 78, και πιο αναλυτικά στη μέση συνολική δαπάνη αυτής της περιόδου. Από το γράφημα 1, φαίνεται πως οι πελάτες που έκανα αγορά πρόσφατα, και αυτοί που αγόρασαν συχνά στο πρώτο μισό της περιόδου εξέτασης, ξανά-αγόρασαν από την επιχείρηση CDNOW στο επόμενο μισό της περιόδου, επιβεβαιώνοντας το μοντέλο RFM. Επίσης, όσο πιο πρόσφατη ήταν η τελευταία αγορά, και όσο πιο συχνές αγορές έκαναν, τόσο πιο μεγάλη ήταν η συνολική μέση δαπάνη τους.



Γράφημα 1: Μέση Συνολική Δαπάνη Εβδομάδων 40-78 ανά Επικαιρότητα και Συχνότητα των Εβδομάδων 1-39 (Fader, Hardie, & Lee, 2005)

Οι Rust και Verhoef (2005) έλαβαν στοιχεία από ένα πάροχο χρηματοοικονομικών υπηρεσιών που ειδικεύεται στις ασφάλειες. Τα δεδομένα αποτελούνται από τις αγορές ασφαλίσεων 1.580 πελατών κατά τη διάρκεια μιας διετίας.

Πίνακας 3: Ανταπόκριση των μοντέλων τμηματοποίησης στο μάρκετινγκ (Rust & Verhoef, 2005)

Μοντέλο	Τμηματοποίηση	Ανταπόκριση σε Ενέργειες AM (%)
Δημογραφικά	Χαμηλό Εισόδημα/Νέοι	7.09
	Χαμηλό Εισόδημα/Ηλικιωμένοι	12.53
	Υψηλό Εισόδημα/Νέοι	26.57
	Υψηλό Εισόδημα/Ηλικιωμένοι	20.00
RFM	Χαμηλή δαπάνη	12.48
	Υψηλή Δαπάνη/παρέκκλιση/σπάνια	13.32
	Υψηλή δαπάνη/πρόσφατη/σπάνια	22.73
	Υψηλή δαπάνη/παρέκκλιση/συχνά	29.78
	Υψηλή δαπάνη/πρόσφατη/συχνά	16.10

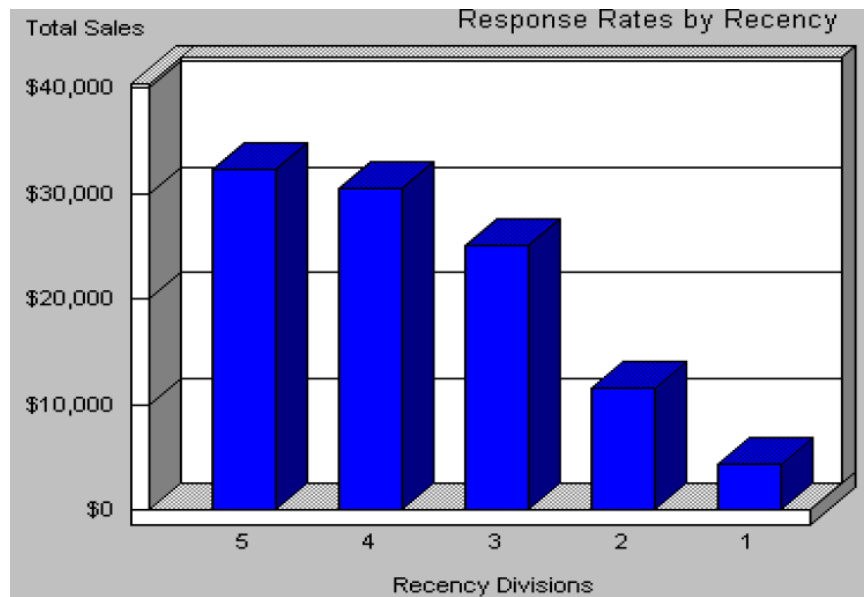
Όπως φαίνεται από τον πίνακα 3, με τη χρήση του μοντέλου RFM, οι πελάτες που ξόδευαν αρκετά σε πακέτα ασφάλειας και αγόραζαν συχνά, άσχετα αν η τελευταία τους αγορά ήταν πρόσφατη ή είχαν μια μικρή παρέκκλιση, είχαν τα καλύτερα ποσοστά ανταπόκρισης σε ενέργειες AM όπως και αναμενόταν από τους ερευνητές Rust και Verhoef (2005). Η πρώτη πρόταση που έκαναν ήταν πως όσο πιο μακροχρόνια η σχέση του πελάτη, όσο μεγαλύτερος αριθμός χρησιμοποιούμενων προϊόντων (αριθμός ασφαλίσεων) και ο μεγάλος αθροιστικός αριθμός αγορών έχουν αρνητικό αντίκτυπο στην ανταπόκριση στις ενέργειες AM, ενώ αν έχει περάσει αρκετά μεγάλο χρονικό διάστημα από την τελευταία αγορά τότε έχει το αναμενόμενο θετικό αντίκτυπο. Για τις ανάγκες της έρευνας δημιούργησαν ένα Ιεραρχικό μοντέλο τμηματοποίησης και παρόλο που είχε πολύ θετικά αποτελέσματα, καταλήγουν πως αυτό το μοντέλο δεν είναι η καλύτερη λύση για όλες τις επιχειρήσεις και το μοντέλο RFM θα ήταν ιδανικότερο.

Ο Hughes (2005) εφάρμοσε το μοντέλο RFM σε 45.000 πελάτες, εταιρείας πώλησης προσωποποιημένων προϊόντων μέσω AM (Άμεσου Μάρκετινγκ) και παρουσίασε τα κάτωθι ευρήματα.



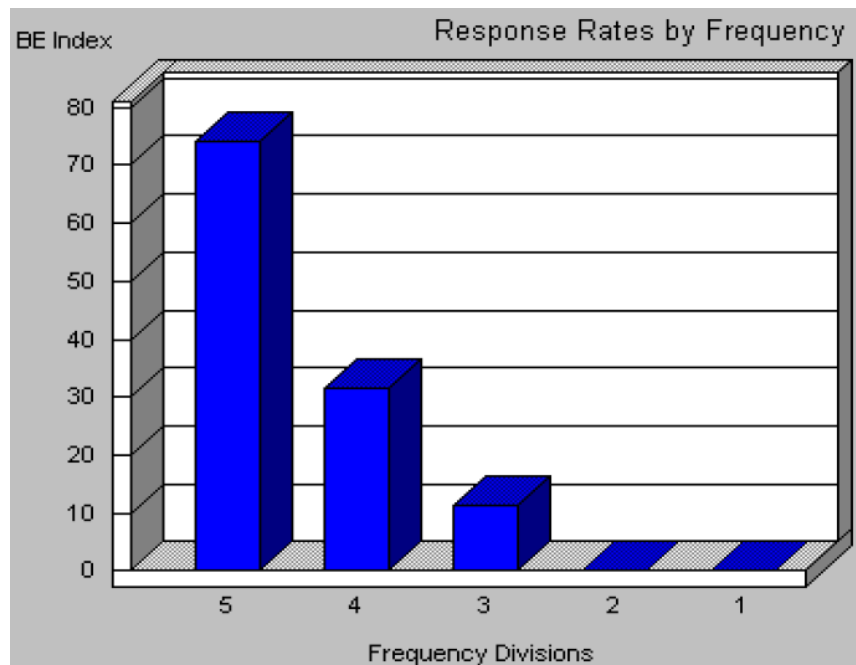
Γράφημα 2: Ποσοστό ανταπόκρισης ανά επικαιρότητα αγοράς (Hughes, 2005)

Στο γράφημα 2 φαίνεται πως η ομάδα πελατών που είχε την δεύτερη πιο πρόσφατη αγορά στην επιχείρηση ανταποκρίθηκαν καλύτερα στη ενέργεια AM. Παρ' όλ' αυτά, όπως φαίνεται από το γράφημα 3, η ομάδα πελατών που είχε πραγματοποιήσει την πιο πρόσφατη αγορά, ξόδεψε περισσότερο.

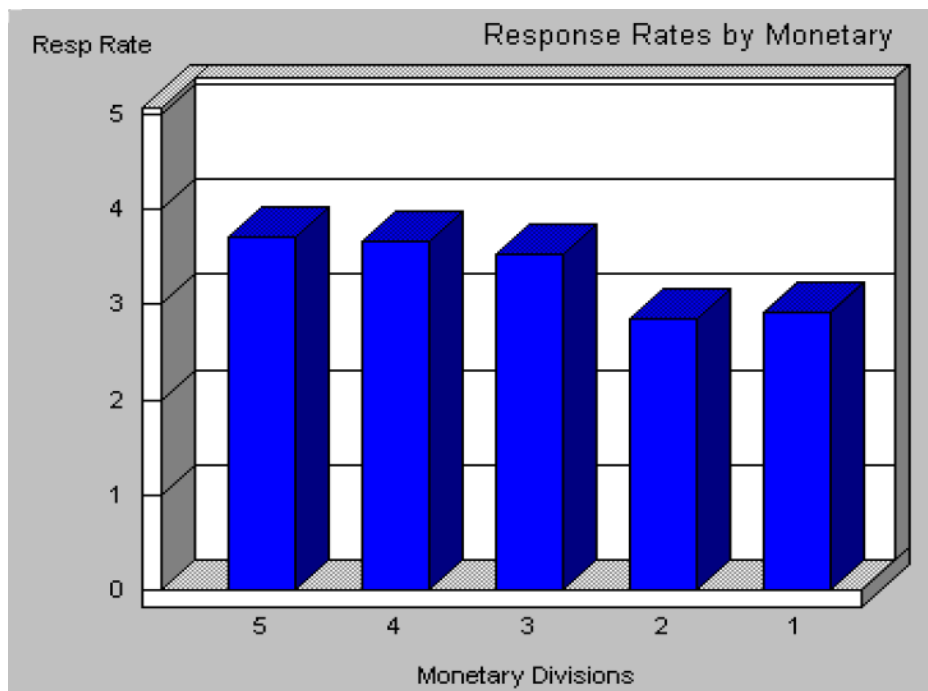


Γράφημα 3: Σύνολο πωλήσεων ανά ομάδα επικαιρότητας αγοράς (Hughes, 2005)

Από το γράφημα 4, φαίνεται πως μόνο οι 3 ομάδες που αγοράζουν πιο συχνά κατάφεραν να αποφέρουν κέρδος στην επιχείρηση.

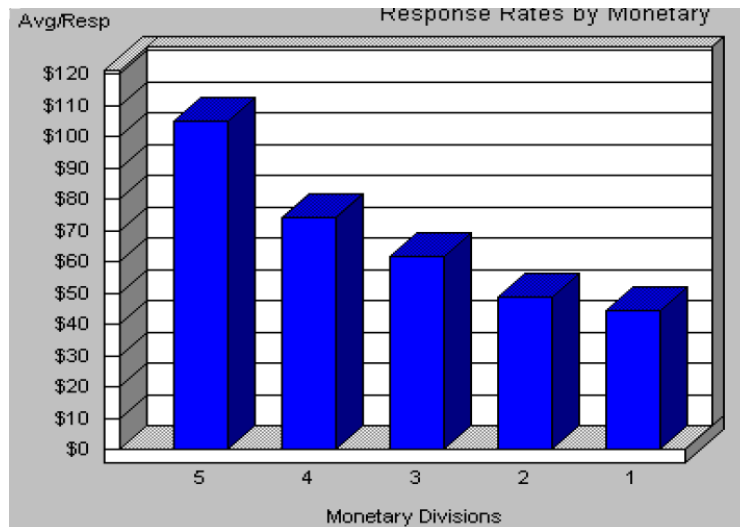


Γράφημα 4: Νεκρό σημείο ανά ομάδα συχνότητας αγορών (Hughes, 2005)

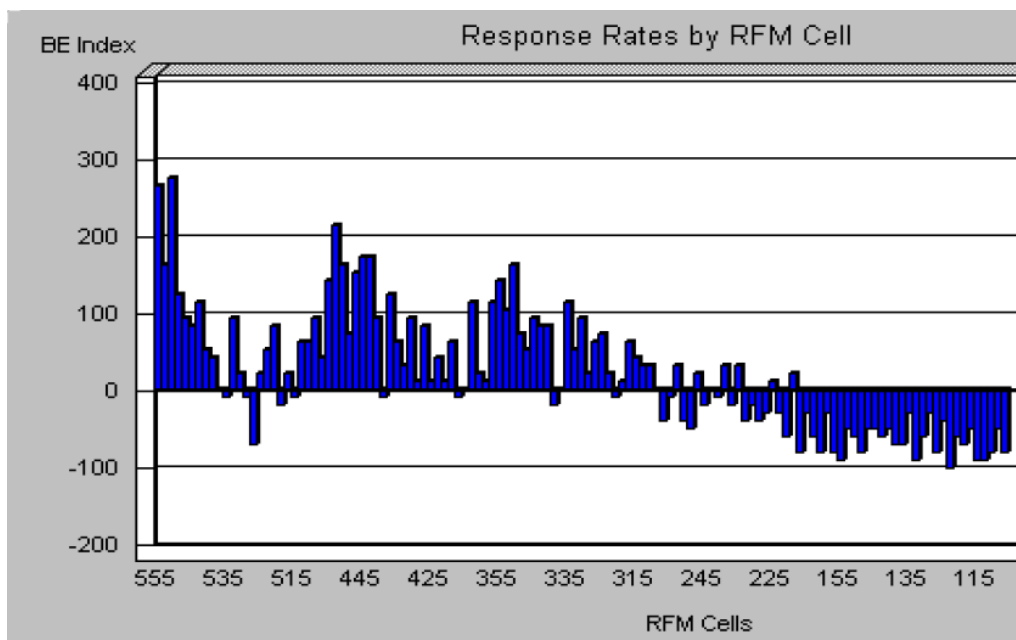


Γράφημα 5: Ρυθμός ανταπόκρισης ανά ομάδα χρηματικής αξίας (Hughes, 2005)

Από το γράφημα 5, φαίνεται πως το ποσό που είχαν ξοδέψει οι πελάτες σε προηγούμενες αγορές, δεν επηρέασε την ανταπόκρισή τους σε αυτή την ενέργεια ΑΜ. Αλλά, επηρέασε την μέση δαπάνη τους στην επιχείρηση, όπως φαίνεται από το γράφημα 6.



Γράφημα 6: Μέση δαπάνη ανά ομάδα χρηματικής αξίας (Hughes, 2005)



Γράφημα 7: Νεκρό σημείο ανά κελί RFM (Hughes, 2005)

Από το γράφημα 7 φαίνεται ότι οι πελάτες με την πιο πρόσφατη αγορά, τις πιο συχνές αγορές, και τη μεγαλύτερη χρηματικά αξία, αποφέρουν το μέγιστο κέρδος στην επιχείρηση.

Οι Zhang et al. (2014) επισημαίνουν την ανεπάρκεια της παραδοσιακής ανάλυσης RFM και προτείνουν στην έρευνα τους την προσθήκη μιας ακόμα διάστασης, της C, όπου πηγάζει από την CLV (Customer Lifetime Value – Αξία χρόνου ζωής του πελάτη). Η C (clumpiness – συσσωμάτωση) υποδηλώνει μια μη σταθερή τάση, συγκεκριμένα προσωρινές αυξήσεις της τάσης - δηλαδή περιόδους κατά τις οποίες ένα γεγονός είναι πιο πιθανό να συμβεί από το μέσο επίπεδο (Zhang, Bradlow, & Small, 2013). Το κύριο εμπειρικό τους εύρημα είναι ότι η C προσθέτει επιπλέον αξία στην προγνωστική δύναμη της RFM, τόσο στα τμήματα της απομάκρυνσης, της συχνότητας εμφάνισης και της χρηματικής αξίας της CLV. Το βασικό αποτέλεσμα της έρευνάς είναι ότι, εκτός από την RFM, η συσσωμάτωση (C) αποτελεί επίσης σημαντικό παράγοντα για τη δημιουργία προφίλ των πελατών και την εκτίμηση του CLV. Το καταδεικνύουν με τη διεξαγωγή μιας λεπτομερούς εμπειρικής μελέτης ενός συνόλου δεδομένων από ένα μεγάλο λιανοπωλητή της Βόρειας Αμερικής. Από μεθοδολογική άποψη, δείχνουν ότι, παρόλο που τα μοντέλα που βασίζονται στο RFM είναι σε θέση να παρέχουν καλές συνολικές εκτιμήσεις της συχνότητας εντός και εκτός δείγματος σε ορισμένες περιπτώσεις, θα οδηγήσουν σε σημαντικά σφάλματα πρόβλεψης σε ατομικό επίπεδο, εάν η συσσωμάτωση υπάρχει αλλά δεν παίρνεται υπόψη. Επεκτείνουν επίσης την ανάλυση σε μια ποικιλία δεδομένων από έξι εταιρείες: δύο παραδοσιακές και τέσσερις διαδικτυακές. Αυτό αποδεικνύει ότι το φαινόμενο συσσωμάτωσης - clumpiness είναι ευρέως διαδεδομένο στο διαδίκτυο ή τουλάχιστον αξίζει να διερευνηθεί. Από τη σκοπιά της διαχειριστικής εφαρμογής μάρκετινγκ, οι πελάτες που παρουσιάζουν συσσωμάτωση σημαίνουν υψηλά πιθανά κέρδη, αλλά και υψηλούς κινδύνους. Μέσω της σωστής διαχείρισης και της σωστής προώθησης, μπορούν να οδηγήσουν σε τεράστια αξία για την εταιρεία με τρόπο που θα μπορούσε να διαφύγει διαφορετικά.

Εμπνευσμένοι από την πρόκληση της επίλυσης των ζητημάτων της CLV, και της σύνοψης των δεδομένων πελατών σε χρήσιμες πληροφορίες για τους διευθυντές μάρκετινγκ, οι Heldt et al. (2021), προτείνουν μια νέα προσέγγιση για την πρόβλεψη της αξίας των πελατών με βάση ένα μοντέλο RFM ανά προϊόν (RFM/P). Αυτή η εναλλακτική λύση συνιστά την ενσωμάτωση των ενεργειών μάρκετινγκ προϊόντων και πελατών συνδυάζοντάς τες για να παρέχουν μια πληρέστερη επισκόπηση της μελλοντικής ταμειακής ροής μιας επιχείρησης. Σε αυτό το μοντέλο, οι αξίες των πελατών εκτιμώνται πρώτα για κάθε προϊόν (ή κατηγορία προϊόντων) και στη συνέχεια αθροίζονται για να προκύψει η συνολική αξία των πελατών. Με αυτόν τον τρόπο, δεν υπάρχει ανάγκη επιλογής μεταξύ της οπτικής των προϊόντων και της οπτικής του πελάτη.

Εφάρμοσαν το μοντέλο τους σε αγοραστικά δεδομένα πελατών από 2 εταιρείες, ένα σούπερ μάρκετ και μία χρηματοοικονομική εταιρεία. Τα αποτελέσματα της μελέτης καταδεικνύουν ότι τα δεδομένα προϊόντων μπορούν να προσθέσουν χρήσιμες πληροφορίες για τη διαχείριση των πόρων μάρκετινγκ και την ακριβέστερη εκτίμηση της CLV. Επιπλέον, μπορούν να μειώσουν το σφάλμα πρόβλεψης της αξίας της πελατειακής βάσης, να βελτιώσουν τα σφάλματα πρόβλεψης της αξίας των μεμονωμένων πελατών και να βοηθήσουν τις εταιρείες να διαχειριστούν καλύτερα την πελατειακή τους βάση. Υπάρχουν ενδείξεις ότι το μοντέλο RFM/P μπορεί να εκτιμήσει την CLV

με μεγαλύτερη ακρίβεια από τα παραδοσιακά συγκεντρωτικά μοντέλα RFM, αποδίδοντας καλύτερα ή τουλάχιστον ισοδύναμα με αυτά. Συνεπώς, οι Heldt et al. (2021) συνηγορούν υπέρ της χρήσης του RFM/P για την πρόβλεψη της αξίας των πελατών. Επιπλέον, το διαχωρισμένο μοντέλο RFM/P δίνει τη δυνατότητα στους διευθυντές να αποκτήσουν μια πιο ολοκληρωμένη στρατηγική εικόνα του μείγματος προϊόντων και του χαρτοφυλακίου πελατών της εταιρείας. Το προτεινόμενο μοντέλο παρέχει τον προσδιορισμό των προϊόντων που αφορούν τους πιο πολύτιμους πελάτες και των πελατών που αγοράζουν τα πιο κερδοφόρα προϊόντα.

3 Μεθοδολογία Έρευνας

Στο τρίτο κεφάλαιο παρουσιάζεται η μεθοδολογία της έρευνας. Αρχικά, θα παρουσιαστεί η Επιστήμη του Σχεδιασμού και η καινοτομία αυτής της έρευνας, και δεύτερον θα παρουσιαστεί η μεθοδολογία Εξόρυξης Δεδομένων. Τέλος, θα γίνει παρουσίαση των δεδομένων που έχουν χορηγηθεί από την εταιρεία.

3.1 Ερευνητική Μεθοδολογία

Η μεθοδολογία που θα ακολουθηθεί είναι πρωτίστως η Εξόρυξη Δεδομένων (Data Mining) για της ανάγκες της ανάλυσης των δεδομένων των 5.000 πελατών και της τμηματοποίησης τους με χρήση της ανάλυσης RFM. Δεύτερον, με βοήθεια της Επιστήμης του Σχεδιασμού, γίνεται προσπάθεια να εφευρεθεί ένα νέο τεχνούργημα που θα δώσει επιπλέον αξία στην τμηματοποίηση της RFM, εισάγοντας μια καινούργια διάσταση.

3.1.1 Design science – Επιστήμη του Σχεδιασμού

Η Επιστήμη του Σχεδιασμού έχει τις ρίζες της στη μηχανική και στις επιστήμες του τεχνητού. Είναι βασικά μια ιδεολογική δομή επίλυσης προβλημάτων. Επιδιώκει τη δημιουργία καινοτομιών που καθορίζουν τις ιδέες, τις πρακτικές, τις τεχνικές δυνατότητες και τα προϊόντα μέσω των οποίων μπορεί να επιτευχθεί αποτελεσματικά και αποδοτικά η ανάλυση, ο σχεδιασμός, η υλοποίηση, η διαχείριση και η χρήση των πληροφοριακών συστημάτων (Hevner et al. 2004).

Η επιστήμη του σχεδιασμού είναι η επιστημονική μελέτη και δημιουργία αντικειμένων-τεχνουργημάτων όπως αυτά αναπτύσσονται και χρησιμοποιούνται από τους ανθρώπους με στόχο την επίλυση πρακτικών προβλημάτων γενικού ενδιαφέροντος (Johannesson & Perjons, 2014). Η ιδεολογική δομή της επιστήμης του σχεδιασμού επιδιώκει να επεκτείνει τα όρια των ανθρώπινων και οργανωτικών δυνατοτήτων δημιουργώντας νέα και καινοτόμα τεχνουργήματα (Hevner et al. 2004).

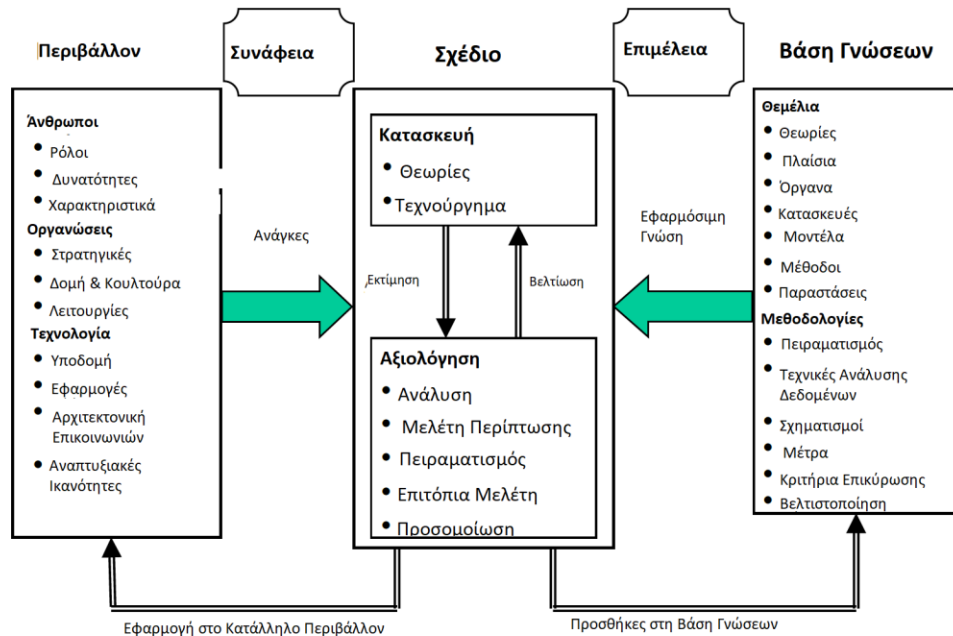
Τα αποτελέσματα της DS περιλαμβάνουν τόσο τα πρόσφατα σχεδιασμένα τεχνουργήματα όσο και τη γνώση σχεδιασμού (DK – Design Knowledge) που παρέχει μια πληρέστερη κατανόηση μέσω θεωριών σχεδιασμού στο γιατί τα τεχνουργήματα ενισχύουν (ή διαταράσσουν) τα σχετικά πλαίσια εφαρμογής (Brocke et al. 2020).

3.1.1.1 Μεθοδολογία Επιστήμης Σχεδιασμού

Η διαδικασία σχεδιασμού είναι μια ακολουθία δραστηριοτήτων εμπειρογνομώνων που παράγει ένα καινοτόμο προϊόν (δηλαδή το τεχνούργημα του σχεδιασμού). Η αξιολόγηση του τεχνουργήματος παρέχει στη συνέχεια πληροφορίες ανατροφοδότησης και καλύτερη κατανόηση του προβλήματος, προκειμένου να βελτιωθεί τόσο η ποιότητα του προϊόντος όσο και η διαδικασία

σχεδιασμού. Αυτός ο κύκλος κατασκευής και αξιολόγησης επαναλαμβάνεται συνήθως αρκετές φορές πριν από τη δημιουργία του τελικού τεχνουργήματος σχεδιασμού. Κατά τη διάρκεια αυτής της δημιουργικής διαδικασίας, ο ερευνητής της επιστήμης του σχεδιασμού πρέπει να έχει επίγνωση της εξέλιξης τόσο της διαδικασίας σχεδιασμού όσο και του τεχνουργήματος σχεδιασμού ως μέρος της έρευνας (Hevner et al. 2004).

Οι March και Smith (1995) προσδιορίζουν δύο διαδικασίες σχεδιασμού και τέσσερα τεχνουργήματα σχεδιασμού που παράγονται από την έρευνα της επιστήμης του σχεδιασμού στα πληροφοριακά συστήματα. Οι δύο διαδικασίες είναι η κατασκευή και η αξιολόγηση. Τα τεχνουργήματα είναι κατασκευές, μοντέλα, μέθοδοι και αναπαραστάσεις. Τα σκόπιμα τεχνουργήματα κατασκευάζονται για να αντιμετωπίσουν μέχρι σήμερα άλυτα προβλήματα. Αξιολογούνται σε σχέση με τη χρησιμότητα που παρέχουν στην επίλυση αυτών των προβλημάτων. Οι κατασκευές παρέχουν τη γλώσσα στην οποία ορίζονται και επικοινωνούνται τα προβλήματα και οι λύσεις. Τα μοντέλα χρησιμοποιούν κατασκευές για να αναπαραστήσουν μια κατάσταση του πραγματικού κόσμου, το σχεδιαστικό πρόβλημα και το χώρο λύσεων του. Τα μοντέλα βοηθούν στην κατανόηση του προβλήματος και της λύσης και συχνά αντιπροσωπεύουν τη σύνδεση μεταξύ των στοιχείων του προβλήματος και της λύσης, επιτρέποντας τη διερεύνηση των επιπτώσεων των αποφάσεων σχεδιασμού και των αλλαγών στον πραγματικό κόσμο. Οι μέθοδοι καθορίζουν τις διαδικασίες. Παρέχουν καθοδήγηση για τον τρόπο επίλυσης των προβλημάτων, δηλαδή για τον τρόπο αναζήτησης του χώρου λύσεων. Αυτές μπορεί να κυμαίνονται από επίσημους, μαθηματικούς αλγορίθμους που ορίζουν ρητά τη διαδικασία αναζήτησης έως ανεπίσημες, κειμενικές περιγραφές προσεγγίσεων "βέλτιστης πρακτικής", ή κάποιο συνδυασμό αυτών. Οι υλοποιήσεις δείχνουν ότι οι κατασκευές, τα μοντέλα ή οι μέθοδοι μπορούν να εφαρμοστούν σε ένα λειτουργικό σύστημα. Επιδεικνύουν τη σκοπιμότητα, επιτρέποντας τη συγκεκριμένη αξιολόγηση της καταλληλότητας ενός τεχνουργήματος για τον επιδιωκόμενο σκοπό του. Επιτρέπουν επίσης στους ερευνητές να μάθουν για τον πραγματικό κόσμο, τον τρόπο με τον οποίο το τεχνουργήμα τον επηρεάζει και τον τρόπο με τον οποίο οι χρήστες το οικειοποιούνται.



Εικόνα 2: Πλαίσιο έρευνας της Επιστήμης του Σχεδιασμού (Hevner, March, Park, & Ram, 2004)

Στην εικόνα 2, οι Hevner et al. (2004) παρουσιάζουν το πλαίσιο έρευνας της επιστήμης του σχεδιασμού. Το περιβάλλον ορίζει το χώρο του προβλήματος στον οποίο βρίσκονται τα φαινόμενα που μας ενδιαφέρουν. Το περιβάλλον αποτελείται από ανθρώπους, (επιχειρηματικούς) οργανισμούς και τις υπάρχουσες ή σχεδιαζόμενες τεχνολογίες τους. Σε αυτό υπάρχουν οι στόχοι, τα καθήκοντα, τα προβλήματα και οι ευκαιρίες που καθορίζουν τις επιχειρηματικές ανάγκες, όπως αυτές γίνονται αντιληπτές από τους ανθρώπους μέσα στον οργανισμό. Οι αντιλήψεις αυτές διαμορφώνονται από τους ρόλους, τις ικανότητες και τα χαρακτηριστικά των ανθρώπων εντός του οργανισμού. Οι επιχειρηματικές ανάγκες εκτιμώνται και αξιολογούνται στο πλαίσιο των οργανωτικών στρατηγικών, της δομής, της κουλτούρας και των υφιστάμενων επιχειρηματικών διαδικασιών. Τοποθετούνται σε σχέση με την υφιστάμενη τεχνολογική υποδομή, τις εφαρμογές, τις αρχιτεκτονικές επικοινωνίας, και τις ικανότητες που προϋπάρχουν. Αυτά καθορίζουν την επιχειρηματική ανάγκη ή το "πρόβλημα" όπως το αντιλαμβάνεται ο ερευνητής. Η διαμόρφωση των ερευνητικών δραστηριοτήτων για την αντιμετώπιση των επιχειρηματικών προβλημάτων εξασφαλίζει τη συνάφεια της έρευνας. Η βάση γνώσεων παρέχει τις πρώτες ύλες από τις οποίες και μέσω των οποίων πραγματοποιείται η έρευνα. Η βάση γνώσεων αποτελείται από θεμέλια και μεθοδολογίες. Προηγούμενες έρευνες και αποτελέσματα από κλάδους αναφοράς παρέχουν θεμελιακές θεωρίες, πλαίσια και μέσα, κατασκευές, μοντέλα, μεθόδους και υλοποιήσεις, που χρησιμοποιούνται στη φάση ανάπτυξης/κατασκευής μιας εμπειρικής μελέτης. Οι μεθοδολογίες παρέχουν κατευθυντήριες γραμμές που χρησιμοποιούνται στη φάση της αιτιολόγησης/αξιολόγησης. Η αυστηρότητα επιτυγχάνεται με την κατάλληλη εφαρμογή των υφιστάμενων βάσεων και μεθοδολογιών.

3.1.1.2 Καινοτομία – Τεχνούργημα Παρούσας Έρευνας

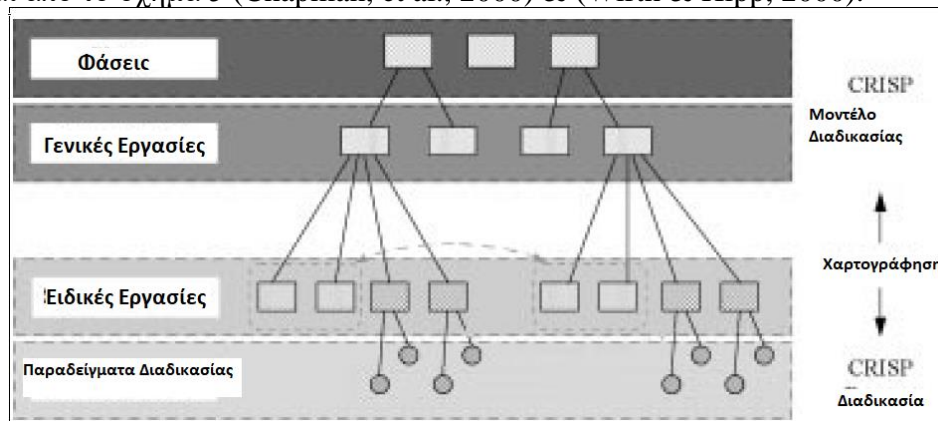
Το τεχνούργημα που θα δημιουργηθεί για τις ανάγκες αυτής της έρευνας θα είναι η προσθήκη μιας ακόμα διάστασης στην ανάλυση RFM. Η διάσταση που θα δημιουργηθεί είναι αυτή της “V”, που πηγάζει από την Αγγλική λέξη “Variety” ή στα Ελληνικά, ποικιλία. Με την εισαγωγή της “V”, η διαδικασία συσταδοποίησης των πελατών θα πάρει υπόψη και την ποικιλία προϊόντων που έχουν αγοράσει. Στα δεδομένα που έχουν δοθεί από την επιχείρηση υπάρχει και αρχείο με τα προϊόντα που αγόρασαν οι πελάτες, μαζί με την υποκατηγορία, την κατηγορία και τον τομέα που ανήκει το κάθε προϊόν. Αρχικά, θα γίνει συλλογή όλων των προϊόντων που έχει ο αγοράσει ο πελάτης, στη συνέχεια θα δημιουργηθεί νέα κολώνα δεδομένων όπου το κάθε προϊόν θα έχει νέα ονομασία η οποία θα περιλαμβάνει την υποκατηγορία, την κατηγορία αλλά και τον τομέα, και τέλος, από αυτή τη κολώνα θα κρατηθούν μόνο οι ξεχωριστές τιμές προϊόντων για το κάθε πελάτη, δηλαδή θα διαγραφούν διπλές και τριπλές τιμές, ώστε να εμφανισθεί η ποικιλία των προϊόντων που έχει αγοράσει. Η διαδικασία της επέκτασης “V” θα αναλυθεί στο επόμενο κεφάλαιο της εμπειρικής μελέτης.

3.1.2 CRISP-DM - Εξόρυξη Δεδομένων

Το CRISP-DM σχεδιάστηκε στα τέλη του 1996, από την ISL (SPSS), την Daimler Chrysler και την NCR, όπου μαζί με τη βοήθεια άλλων δημιούργησαν την κοινοπραξία CRISP-DM (CRoss-Industry Standard Process for Data Mining) όπου χρηματοδοτήθηκε από την Ευρωπαϊκή Επιτροπή (Chapman, et al., 2000).

3.1.2.1 Μεθοδολογία

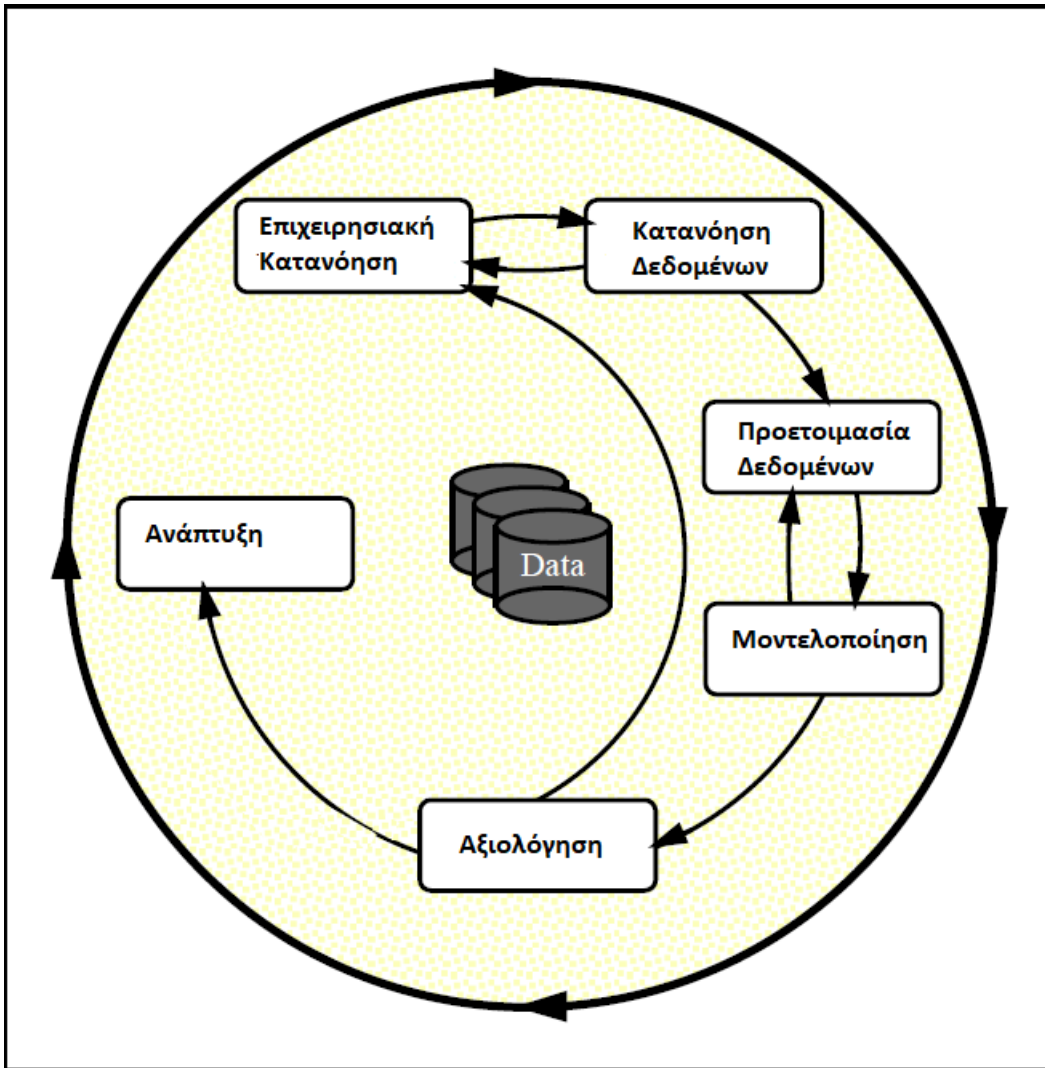
Η μεθοδολογία εξόρυξης δεδομένων CRISP-DM περιγράφεται με όρους ενός ιεραρχικού μοντέλου διαδικασιών, το οποίο αποτελείται από σύνολα εργασιών που περιγράφονται σε τέσσερα επίπεδα αφαίρεσης (από το γενικό έως το ειδικό επίπεδο). Πρώτα είναι η φάσεις, μετά οι γενικές εργασίες, στη συνέχεια οι εξειδικευμένες εργασίες και τέλος τα παραδείγματα διαδικασιών, όπως φαίνεται και από το σχήμα 5 (Chapman, et al., 2000) & (Wirth & Hipp, 2000).



Σχήμα 5: Ανάλυση της μεθοδολογίας CRISP-DM σε τέσσερα επίπεδα (Chapman, et al., 2000)

Στο ανώτερο επίπεδο, η διαδικασία εξόρυξης δεδομένων οργανώνεται σε διάφορες φάσεις - κάθε φάση αποτελείται από διάφορες γενικές εργασίες δευτέρου επιπέδου. Αυτό το δεύτερο επίπεδο καλείται γενικό, επειδή επιδιώκεται να είναι αρκετά γενικό ώστε να καλύπτει όλες τις πιθανές καταστάσεις εξόρυξης δεδομένων. Οι γενικές εργασίες επιδιώκεται να είναι όσο το δυνατόν πληρέστερες και σταθερότερες. Πλήρης σημαίνει ότι καλύπτει τόσο την όλη διαδικασία εξόρυξης δεδομένων όσο και όλες τις πιθανές εφαρμογές εξόρυξης δεδομένων. Το τρίτο επίπεδο, το επίπεδο εξειδικευμένων εργασιών, είναι το μέρος όπου περιγράφεται ο τρόπος με τον οποίο οι ενέργειες των γενικών εργασιών πρέπει να εκτελούνται σε ορισμένες ειδικές καταστάσεις. Για παράδειγμα, στο δεύτερο επίπεδο μπορεί να υπάρχει μια γενική εργασία που ονομάζεται καθαρισμός δεδομένων. Το τρίτο επίπεδο περιγράφει πώς διαφέρει αυτή η εργασία σε διαφορετικές καταστάσεις, όπως ο καθαρισμός αριθμητικών τιμών σε σχέση με τον καθαρισμό κατηγορικών τιμών ή αν ο τύπος του προβλήματος είναι ομαδοποίηση ή προγνωστική μοντελοποίηση. Η περιγραφή των φάσεων και των εργασιών ως διακριτά βήματα που εκτελούνται με συγκεκριμένη σειρά αντιπροσωπεύει μια εξιδανικευμένη ακολουθία γεγονότων. Στην πράξη, πολλές από τις εργασίες μπορούν να εκτελεστούν με διαφορετική σειρά και συχνά είναι απαραίτητη η επιστροφή επανειλημμένα σε προηγούμενες εργασίες και να επαναληφθούν ορισμένες ενέργειες. Το μοντέλο διεργασιών δεν επιχειρεί να συλλάβει όλες αυτές τις πιθανές διαδρομές μέσω της διαδικασίας εξόρυξης δεδομένων, διότι αυτό θα απαιτούσε ένα υπερβολικά πολύπλοκο μοντέλο διαδικασίας. Το τέταρτο επίπεδο, το παράδειγμα διαδικασίας, είναι μια καταγραφή των ενεργειών, των αποφάσεων και των αποτελεσμάτων μιας πραγματικής δέσμευσης εξόρυξης δεδομένων. Μια περίπτωση διαδικασίας οργανώνεται σύμφωνα με τα καθήκοντα που ορίζονται στα υψηλότερα επίπεδα, αλλά αντιπροσωπεύει αυτό που πραγματικά συνέβη σε μια συγκεκριμένη εργασία, και όχι τι συμβαίνει γενικά (Charman, et al., 2000).

Το μοντέλο αναφοράς CRISP-DM για την εξόρυξη δεδομένων παρέχει μια επισκόπηση του κύκλου ζωής ενός έργου εξόρυξης δεδομένων. Περιέχει τις φάσεις ενός έργου, τις αντίστοιχες εργασίες τους και τα αποτελέσματά τους. Ο κύκλος ζωής ενός έργου εξόρυξης δεδομένων αναλύεται σε έξι φάσεις οι οποίες παρουσιάζονται στο Σχήμα 6. Η αλληλουχία των φάσεων δεν είναι αυστηρή. Τα βέλη υποδεικνύουν μόνο τις πιο σημαντικές και συχνές εξαρτήσεις μεταξύ των φάσεων, αλλά σε ένα συγκεκριμένο έργο, εξαρτάται από το αποτέλεσμα κάθε φάσης, για το ποια φάση, ή ποια συγκεκριμένη εργασία μιας φάσης, πρέπει να εκτελεστεί στη συνέχεια (Wirth & Hipp, 2000).



Σχήμα 6: Φάσεις του τρέχοντος μοντέλου διαδικασιών CRISP-DM για την εξόρυξη δεδομένων (Wirth & Hipp, 2000)

Οι Wirth και Hipp (2000) εξηγούν τις διάφορες φάσεις του μοντέλου διαδικασιών της CRISP – DM:

- **Επιχειρησιακή κατανόηση**

Αυτή η αρχική φάση επικεντρώνεται στην κατανόηση των στόχων και των απαιτήσεων του έργου από επιχειρηματική άποψη και, στη συνέχεια, στη μετατροπή αυτής της γνώσης σε ορισμό του προβλήματος εξόρυξης δεδομένων και σε ένα προκαταρκτικό σχέδιο έργου σχεδιασμένο για την επίτευξη των στόχων.

• Κατανόηση δεδομένων

Η φάση της κατανόησης δεδομένων ξεκινά με μια αρχική συλλογή δεδομένων και συνεχίζει με δραστηριότητες προκειμένου να εξοικειωθεί με τα δεδομένα, να εντοπίσει προβλήματα ποιότητας δεδομένων, να ανακαλύψει τις πρώτες γνώσεις για τα δεδομένα ή να εντοπίσει ενδιαφέροντα υποσύνολα για να σχηματίσει υποθέσεις για κρυμμένες πληροφορίες. Υπάρχει στενή σύνδεση μεταξύ της επιχειρησιακής κατανόησης και της κατανόησης των δεδομένων. Η διατύπωση του προβλήματος εξόρυξης δεδομένων και το σχέδιο έργου απαιτούν τουλάχιστον κάποια κατανόηση των διαθέσιμων δεδομένων.

• Προετοιμασία δεδομένων

Η φάση της προετοιμασίας δεδομένων καλύπτει όλες τις δραστηριότητες για την κατασκευή του τελικού συνόλου δεδομένων (δεδομένα που θα τροφοδοτηθούν στα εργαλεία μοντελοποίησης) από τα αρχικά ακατέργαστα δεδομένα. Οι εργασίες προετοιμασίας δεδομένων είναι πιθανό να εκτελούνται πολλές φορές και όχι με προκαθορισμένη σειρά. Οι εργασίες περιλαμβάνουν την επιλογή πινάκων, εγγραφών και χαρακτηριστικών, τον καθαρισμό δεδομένων, την κατασκευή νέων χαρακτηριστικών και τον μετασχηματισμό των δεδομένων για τα εργαλεία μοντελοποίησης.

• Μοντελοποίηση

Σε αυτή τη φάση, επιλέγονται και εφαρμόζονται διάφορες τεχνικές μοντελοποίησης και οι παράμετροι τους βαθμονομούνται σε βέλτιστες τιμές. Συνήθως, υπάρχουν διάφορες τεχνικές για τον ίδιο τύπο προβλήματος εξόρυξης δεδομένων. Ορισμένες τεχνικές απαιτούν συγκεκριμένες μορφές δεδομένων. Υπάρχει στενή σύνδεση μεταξύ της προετοιμασίας δεδομένων και της μοντελοποίησης. Συχνά, ένας αναλυτής συνειδητοποιεί προβλήματα δεδομένων κατά τη μοντελοποίηση ή παίρνει ιδέες για την κατασκευή νέων δεδομένων.

• Αξιολόγηση

Σε αυτό το στάδιο του έργου έχουν κατασκευαστεί ένα ή περισσότερα μοντέλα που φαίνεται να έχουν υψηλή ποιότητα, από την άποψη της ανάλυσης δεδομένων. Πριν προχωρήσει η φάση στην τελική ανάπτυξη του μοντέλου, είναι σημαντικό να αξιολογηθεί πιο διεξοδικά το μοντέλο και να επανεξεταστούν τα βήματα που εκτελέστηκαν για την κατασκευή του μοντέλου, ώστε να σιγουρευτεί ότι επιτυγχάνει σωστά τους επιχειρηματικούς στόχους. Ένας βασικός στόχος είναι να προσδιοριστεί εάν υπάρχει κάποιο σημαντικό επιχειρηματικό ζήτημα που δεν έχει ληφθεί επαρκώς υπόψη. Στο τέλος αυτής της φάσης, θα πρέπει να ληφθεί απόφαση σχετικά με τη χρήση των αποτελεσμάτων της εξόρυξης δεδομένων.

• **Ανάπτυξη**

Η δημιουργία του μοντέλου δεν αποτελεί γενικά το τέλος του έργου. Συνήθως, η αποκτηθείσα γνώση θα πρέπει να οργανωθεί και να παρουσιαστεί με τρόπο ώστε ο πελάτης να μπορεί να τη χρησιμοποιήσει. Ανάλογα με τις απαιτήσεις, η φάση ανάπτυξης μπορεί να είναι τόσο απλή όσο η δημιουργία μιας έκθεσης ή τόσο σύνθετη όσο η εφαρμογή μιας επαναλαμβανόμενης διαδικασίας εξόρυξης δεδομένων. Σε πολλές περιπτώσεις θα είναι ο χρήστης, όχι ο αναλυτής δεδομένων, που θα εκτελέσει τα βήματα ανάπτυξης. Σε κάθε περίπτωση, είναι σημαντικό να κατανοηθούν εκ των προτέρων ποιες ενέργειες θα πρέπει να πραγματοποιηθούν προκειμένου να γίνει πραγματική χρήση των μοντέλων που δημιουργήθηκαν.

Στον Πίνακα 4 γίνεται επισκόπηση των εργασιών της CRISP-DM αλλά και των αποτελεσμάτων τους.

Πίνακας 4: Επισκόπηση των εργασιών της CRISP-DM και των αποτελεσμάτων της. (Wirth & Hipp, 2000)

Επιχειρησιακή Κατανόηση	Κατανόηση Δεδομένων	Προετοιμασία Δεδομένων	Μοντελοποίηση	Αξιολόγηση	Ανάπτυξη
<ul style="list-style-type: none"> Καθορισμός επιχειρηματικών στόχων <p>Ιστορικό Επιχειρηματικοί στόχοι Κριτήρια επιχειρηματικής επιτυχίας</p>	<ul style="list-style-type: none"> Συλλογή αρχικών δεδομένων <p>Έκθεση αρχικής συλλογής δεδομένων</p>	<ul style="list-style-type: none"> Σύνολο δεδομένων <p>Περιγραφή συνόλου δεδομένων</p>	<ul style="list-style-type: none"> Επιλογή Τεχνικής μοντελοποίησης <p>Τεχνική μοντελοποίησης Παραδοχές μοντελοποίησης</p>	<ul style="list-style-type: none"> Αξιολόγηση αποτελεσμάτων <p>Αξιολόγηση των αποτελεσμάτων της εξόρυξης δεδομένων σε σχέση με τα επιχειρηματικά κριτήρια επιτυχίας Εγκεκριμένα μοντέλα</p>	<ul style="list-style-type: none"> Ανάπτυξη σχεδίου <p>Σχέδιο ανάπτυξης</p>
<ul style="list-style-type: none"> Αξιολόγηση της κατάστασης <p>Καταγραφή των πόρων</p>	<ul style="list-style-type: none"> Περιγραφή ή δεδομένων <p>Αναφορά περιγραφής δεδομένων</p>	<ul style="list-style-type: none"> Επιλογή δεδομένων <p>Αιτιολόγηση για συμπερίληψη/αποκλεισμό</p>	<ul style="list-style-type: none"> Δημιουργία σχεδίου δοκιμής <p>Δοκιμή σχεδιασμού</p>	<ul style="list-style-type: none"> Διαδικασία αναθεώρησης <p>Αναθεώρηση της διαδικασίας</p>	<ul style="list-style-type: none"> Παρακολούθηση του σχεδίου και Συντήρηση <p>Παρακολούθηση και Σχέδιο συντήρησης</p>

Απαιτήσεις, Υποθέσεις και Περιορισμοί Κίνδυνοι και απρόβλεπτα Ορολογία Κόστος και οφέλη					
<ul style="list-style-type: none"> Καθορισμός των στόχων εξόρυξης δεδομένων <p>Στόχοι εξόρυξης δεδομένων Κριτήρια επιτυχίας της Εξόρυξης Δεδομένων</p>	<ul style="list-style-type: none"> Εξερεύνηση δεδομένων <p>Έκθεση εξερεύνησης δεδομένων</p>	<ul style="list-style-type: none"> Καθαρισμός δεδομένων <p>Έκθεση καθαρισμού δεδομένων</p>	<ul style="list-style-type: none"> Κατασκευή Μοντέλων <p>Ρυθμίσεις παραμέτρων Μοντέλα Περιγραφή μοντέλου</p>	<ul style="list-style-type: none"> Καθορισμός επόμενων βημάτων <p>Κατάλογος πιθανών ενεργειών Απόφαση</p>	<ul style="list-style-type: none"> Παραγωγή τελικής έκθεσης <p>Τελική έκθεση Τελική παρουσίαση</p> <ul style="list-style-type: none"> Αναθεώρηση έργου <p>Τεκμηρίωση εμπειρίας</p>
<ul style="list-style-type: none"> Παραγωγή Σχεδίου Έργου <p>Σχέδιο έργου Αρχική αξιολόγηση εργαλείων και τεχνικών</p>	<ul style="list-style-type: none"> Επαλήθευση της ποιότητας των δεδομένων <p>Αναφορά ποιότητας δεδομένων</p>	<ul style="list-style-type: none"> Κατασκευή δεδομένων <p>Παράγωγα χαρακτηριστικά Παραγόμενες εγγραφές</p> <ul style="list-style-type: none"> Ενσωμάτωση δεδομένων <p>Συγχωνευμένα δεδομένα</p> <ul style="list-style-type: none"> Μορφή δεδομένων <p>Αναδιαμορφωμένα δεδομένα</p>	<ul style="list-style-type: none"> Μοντέλο αξιολόγησης <p>Αξιολόγηση μοντέλου Αναθεωρημένη παράμετρος Ρυθμίσεις</p>		

3.2 Διαθεσιμότητα Δεδομένων

Για την παρούσα έρευνα, χρησιμοποιήθηκαν δεδομένα αγοραστικών καλαθιών 5.000 πελατών εταιρείας λιανικής βελτίωσης σπιτιού & ιδιοκατασκευής μαζί με όλα τα μεμονωμένα προϊόντα που αγόρασε κάθε πελάτης. Οι πελάτες επιλέχθηκαν επειδή πραγματοποίησαν τουλάχιστον μία αγορά κατά τη διάρκεια της περιόδου παρατήρησης (5 Μαρτίου 2012 - 4 Μαρτίου 2014). Επίσης, τα δεδομένα είναι από «Γνωστούς Πελάτες» όπου έγιναν "γνωστοί" όταν παρείχαν έγκυρη ταχυδρομική διεύθυνση. Εκτός από τις αγορές, το σύνολο δεδομένων περιλαμβάνει λεπτομερείς ιεραρχίες προϊόντων και χαρακτηριστικά προϊόντων, πληροφορίες για το κατάστημα και δημογραφικά για τον πελάτη. Η εταιρεία ανήκει στις 500 μεγαλύτερες εταιρείες παγκοσμίως (Fortune 500), και επιθυμεί να κρατήσει την ανωνυμία της.

Τα δεδομένα χωρίζονται σε 3 κύριες κατηγορίες

1. **Πελάτες** - Ένας μεμονωμένος πελάτης της εταιρείας. Για την συγκεκριμένη έρευνα, έχει επιλεγθεί η κατηγορία του νοικοκυριού, ενώ έχουν εξαιρεθεί τυχαίοι πελάτες και επιχειρήσεις (B2B)
2. **Αγορές** - ένα και μόνο προϊόν που αγοράζεται κατά τη διάρκεια μιας και μόνο συναλλαγής
3. **Προϊόντα** – Ένας συγκεκριμένος κωδικός SKU προϊόντος που διατίθεται από την εταιρεία

Και οι 3 κατηγορίες, έχουν την εξής μορφή:

Όνομα Στήλης	Τύπος Δεδομένου	Σημειώσεις SQL	Σημειώσεις / Επεξηγήσεις
Το όνομα της στήλης. Αυτό είναι σχεδόν πάντα όπως εμφανίζεται στο σύστημα της εταιρείας, για ευκολία στη μετάφραση κατά τη διαβίβαση ερωτήσεων/σχολίων στον χορηγό δεδομένων.	Καθορίζει τον τύπο δεδομένων για εύκολη δημιουργία πίνακα, π.χ. "string" ή "integer".	Δηλώνει αν η στήλη είναι πρωτεύον ή/και ξένο κλειδί (Primary and/or Foreign Key)	Μια απλή επεξήγηση κειμένου για το τι σημαίνει η στήλη

3.2.1 Λεδομένα Πελατών

1. Το WCAI-Group έχει τρεις πιθανές τιμές: Known Customers – Γνωστοί Πελάτες (KC), και Randomly Selected Customers - Τυχαία επιλεγμένοι πελάτες (ALL). Οι πελάτες αυτοί επιλέχθηκαν επειδή πραγματοποίησαν τουλάχιστον μία αγορά κατά τη διάρκεια της περιόδου παρατήρησης (5 Μαρτίου 2012 - 4 Μαρτίου 2014). Οι Γνωστοί Καταναλωτές έγιναν "γνωστοί" όταν παρείχαν έγκυρη ταχυδρομική διεύθυνση, ενώ οι Γνωστοί Επαγγελματίες είναι εργολάβοι και άλλοι ειδικοί στον τομέα των κατασκευών που έχουν ανοίξει επαγγελματικό λογαριασμό στην εταιρεία. Οι τυχαία επιλεγμένοι πελάτες είναι ένα πραγματικό τυχαίο δείγμα όλων των πελατών που πραγματοποίησαν τουλάχιστον μία αγορά κατά τη διάρκεια του παραθύρου παρατήρησης. Ως εκ τούτου, περιέχει ορισμένους Γνωστούς Επαγγελματίες, ορισμένους Γνωστούς Καταναλωτές και ορισμένους άγνωστους καταναλωτές. Στην παρούσα έρευνα, έχουν κρατηθεί μόνο οι Γνωστοί Πελάτες, και συγκεκριμένα, οι ιδιώτες. Έχουν εξαιρεθεί επαγγελματίες και άγνωστοι πελάτες.
2. Ο "Input Customer" που αναφέρεται σε αυτόν τον πίνακα αναφέρεται σε ένα μόνο μέλος του νοικοκυριού, συνήθως όποιο όνομα εμφανίζεται πρώτο στον φάκελο πελάτη. Ο "Input Customer" μπορεί να είναι ή να μην είναι ο επικεφαλής του νοικοκυριού.
3. Το συνολικό φύλο του νοικοκυριού καθορίζεται από την παρουσία του "M" ή/και του "F" στις στήλες Gender-1st, Gender-2nd και Gender. Έτσι, εάν δεν υπάρχει "F", αλλά υπάρχει τουλάχιστον ένα "M", τότε πρόκειται για νοικοκυριό μόνο από άνδρες. Εάν υπάρχει οποιοσδήποτε συνδυασμός M/F, πρόκειται για νοικοκυριό ανδρών και γυναικών. Εάν δεν υπάρχει ούτε "M" ούτε "F", τότε το φύλο του νοικοκυριού είναι άγνωστο.
4. Η εταιρεία διαχωρίζει τους πελάτες σε επαγγελματίες και καταναλωτές (ιδιώτες). Τιμή 2 = "Professional" και 3 = "Customer". Ενώ μπορεί να υπάρχουν κάποια δημογραφικά δεδομένα, όπως πληροφορίες για το νοικοκυριό και πληροφορίες για την τοποθεσία για ένα "2", τα δεδομένα αυτά μπορεί να αναφέρονται στην επιχείρηση του πελάτη ή μπορεί να αναφέρονται στο σπίτι του πελάτη και η WCAI δεν έχει τρόπο να διακρίνει αυτό το στοιχείο για τους επαγγελματίες.
5. Η εταιρεία λαμβάνει ενημερωμένα δημογραφικά στοιχεία για τους πελάτες της σε μηνιαία βάση. Τα δεδομένα εδώ αντλήθηκαν στις 11 Φεβρουαρίου 2014 και ήταν επίκαιρα μέσα σε αυτόν τον μήνα.
6. Η καθαρή θέση των νοικοκυριών (House Equity) είναι μια εκτίμηση που παρέχεται τόσο σε ακατέργαστα δολάρια όσο και σε έναν κωδικό. Μόνο ένα μικρό ποσοστό των νοικοκυριών στο σύνολο των δεδομένων έχει διαθέσιμες αυτές τις πληροφορίες. Ωστόσο,

όταν υπάρχουν δεδομένα, έρευνες της εταιρείας δείχνουν ότι τα δεδομένα δεν έχουν συνάφεια. Οι τιμές είναι ασυνεπείς και έχουν μεγάλα και ποικίλα εύρη. Η εταιρεία προτείνει να αγνοηθούν εντελώς αυτά τα δεδομένα ως αναξιόπιστα.

7. Οι στήλες που σημειώνονται με αστερίσκο (*) έχουν πίνακες διανομής.
8. Τα πεδία που είναι NULL ή BLANK (κενά) θα πρέπει να θεωρούνται "άγνωστα". Δεν υπάρχει καμία διαφοροποίηση μεταξύ των δύο.

Στον πίνακα 5 γίνεται παρουσίαση των δεδομένων των πελατών, μαζί με σημειώσεις για το κάθε στοιχείο.

Πίνακας 5: Δεδομένων Πελατών

Όνομα Στήλης	Τύπος Δεδομένου	Σημειώσεις SQL	Σημειώσεις / Επεξηγήσεις
hh_id	String	Primary Key	Μοναδικό αναγνωριστικό για το νοικοκυριό
Wcai-Group*	String		Καθορίζει αν το νοικοκυριό ανήκει στο τμήμα Γνωστός πελάτης (KC), Γνωστός επαγγελματίας (KP) ή τυχαία επιλεγμένο τμήμα (ALL).
Customer-Segment*	String		Ο εσωτερικός προσδιορισμός των πελατών της εταιρείας. Αυτός είναι ο μόνος τρόπος με τον οποίο η εταιρεία τμηματοποιεί τους πελάτες. 2 = επαγγελματίας, 3 = καταναλωτής (ιδιώτης)
Marketing-Flag*	String		Η τιμή "N" σημαίνει ότι η εταιρεία δεν είναι σε θέση να προωθεί ενέργειες μάρκετινγκ σε αυτό το άτομο (N = 0,47%).
Postal-Code	String		Ταχυδρομικός κώδικας κατοικίας του νοικοκυριού
state*	String		Πολιτεία κατοικίας του νοικοκυριού
Age-estimate	Integer		Η εκτιμώμενη ηλικία του πελάτη εισόδου,

Customer-ethnicity*	String		Η εθνικότητα του πελάτη (input customer στην προκειμένη περίπτωση)
gender*	String		Καθορίζει αν ο πελάτης εισόδου είναι άνδρας ή γυναίκα
Gender-1st*	String		Προσδιορίζει αν ένα άλλο μέλος του νοικοκυριού είναι άνδρας ή γυναίκα ή άγνωστο.
Gender-2nd*	String		Προσδιορίζει αν ένα άλλο μέλος του νοικοκυριού είναι άνδρας ή γυναίκα ή άγνωστο.
Economic-index*	String		Το εύρος τιμών προβλέπει πόσο πιθανό είναι ο πελάτης να λάβει έγκριση πίστωσης, 01 = Πιθανότητα οικονομικά σταθερός
Income-high*	String		Υποδεικνύει το εκτιμώμενο εισόδημα σε υψηλότερες κλίμακες. Το εισόδημα είτε δηλώνεται μόνο του είτε εκτιμάται με βάση διάφορους δημογραφικούς παράγοντες
income *	String		Υποδεικνύει το εκτιμώμενο εισόδημα του νοικοκυριού. Είτε αυτοαναφερόμενο είτε εκτιμώμενο
Income-n*	String		Υποδεικνύει το εκτιμώμενο εισόδημα των νοικοκυριών σε στενά όρια
investment *	Integer		Καταμέτρηση του εκτιμώμενου αριθμού επενδυτικών ακινήτων στέγασης εκτός κατοικίας που ανήκουν στο νοικοκυριό
Invest-real-estate *	String		Η σημαία δηλώνει ότι το νοικοκυριό κατέχει ακίνητο εκτός από ακίνητο μόνιμης κατοικίας.

worth*	String		Σύνοψη των δεδομένων που αντιπροσωπεύουν το ενεργητικό μείον το παθητικό
occupation *	String		Ο κωδικός δηλώνει την κατηγορία επαγγέλματος του εισερχόμενου ατόμου
Occupation-detail	String		Ο κωδικός υποδεικνύει το λεπτομερές επάγγελμα του ατόμου (input customer)
Home-Assess-actual	String		Ο κωδικός υποδεικνύει το ποσό στο οποίο η κατοικία αξιολογήθηκε για τελευταία φορά από την τοπική αυτοδιοίκηση.
Home-Asses-range *	String		Ο κωδικός υποδεικνύει το εύρος των αξιών κατοικίας, σε δολάρια, στο οποίο η κατοικία εκτιμήθηκε για τελευταία φορά από την τοπική αυτοδιοίκηση
Home-Equity-actual	String		Ο κωδικός υποδεικνύει το ποσό του διαθέσιμου κεφαλαίου της κατοικίας
Home-Equity-range*	String		Ο κωδικός υποδεικνύει το εκτιμώμενο διαθέσιμο εύρος ιδίων κεφαλαίων για ένα ενυπόθηκο δάνειο
Home-lend-actual	String		Δανειστικά ίδια κεφάλαια, υπολογιζόμενα στο 80% της αγοραίας αξίας μείον το σύνολο των δανείων. Προκύπτει από δημόσια αρχεία
Home-lend-range*	String		Ο κωδικός υποδεικνύει το εκτιμώμενο διαθέσιμο κεφάλαιο για ένα ενυπόθηκο δάνειο
Home-ext*	String		Δείχνει τον τύπο του υλικού που χρησιμοποιήθηκε για το εξωτερικό της κατοικίας

Home-temp-source*	String		Δείχνει τι χρησιμοποιείται για τη θέρμανση του σπιτιού
Home-res-length*	String		Δείχνει το συνολικό χρόνο, σε έτη, που ένα νοικοκυριό ζει στην τρέχουσα διεύθυνση.
Home-loan-original	String		Δείχνει το ποσό του πρώτου δανείου για την κατοικία
Home-loan-ac *	String		Δείχνει το συνολικό ποσό του στεγαστικού δανείου για την κατοικία. Το πεδίο αυτό είναι το σύνολο όλων των δανείων μαζί, εάν υπάρχουν πολλαπλά δάνεια για την κατοικία.
Home-loan-2-ac	String		Υποδεικνύει το ποσό του δεύτερου δανείου για την κατοικία, κατά περίπτωση
Home-loan-3-ac	String		Δείχνει το ποσό του τρίτου δανείου για την κατοικία
Home-loan-2-range*	String		Υποδεικνύει εύρος τιμών για το δεύτερο δάνειο της κατοικίας
Home-loan-3-range*	String`		Υποδεικνύει εύρος τιμών για το τρίτο δάνειο της κατοικίας
Home-loan-value-range *	String		Δείχνει το εύρος τιμών για το σύνολο των δανείων ως ποσοστό της αγοραίας αξίας της κατοικίας, από δημόσια στοιχεία
Home-loan-value-est	String		Εκφράζει το ποσό των συνολικών δανείων (τρεχόντων και παρελθόντων) ως ποσοστό της αγοραίας αξίας της κατοικίας, που προκύπτει από δημόσια στοιχεία
Home-sq-foot-est	String		Εκτιμώμενο μέγεθος του οικοπέδου σε τετραγωνικά μέτρα
Home-sq-foot-range*	String		Δείχνει το μέγεθος του οικοπέδου σε τετραγωνικά μέτρα.

Home-mark-value-est	String		Υποδεικνύει την εκτιμώμενη αγοραία αξία της κατοικίας. Κατασκευάστηκε με πολλαπλές πηγές πηγών και γεωγραφικών δεδομένων σε επίπεδο νοικοκυριού.
Home-own-rent*	String		Δείχνει αν το νοικοκυριό είναι ιδιοκτήτης ή ενοικιαστής της κατοικίας (0 = owner/ιδιοκτήτης)
Home-pool*	String		Δείχνει αν υπάρχει πισίνα στην κατοικία (Y = True/αληθής)
Home-property*	String		Δείχνει τη συγκεκριμένη κατηγορία της κατοικίας (π.χ. μονοκατοικία, πολυκατοικία κ.λπ.)
Home-amount-actual	String		Δείχνει το ποσό που καταβλήθηκε για την αγορά της κατοικίας
Home-amount-range*	String		Δείχνει το ποσό αγοράς της κατοικίας σε κλίμακες
Home-purchase-date	Date		Περιέχει το έτος και τον μήνα κατά τον οποίο αγοράστηκε η κατοικία
Home-purchase-year	Date		Το έτος κατά το οποίο αγοράστηκε η κατοικία
Home-roof*	String		Δείχνει το υλικό από το οποίο είναι κατασκευασμένη η στέγη του σπιτιού
Home-rooms*	Integer		Αριθμός δωματίων στο εσωτερικό της κατοικίας
Home-sq-foot-range*	String		Δείχνει το μέγεθος του σπιτιού σε τετραγωνικά μέτρα, χωρισμένο σε κλίμακες
Home-sq-foot-actual	Integer		Δείχνει το μέγεθος του σπιτιού σε τετραγωνικά μέτρα
Home-year-built-ac	Date		Το έτος κατασκευής της κατοικίας

Home-year-built-range*	Date		Το έτος κατασκευής της κατοικίας σε κλίμακες
Market-val-quality-index*	String		Υποδεικνύει τη μέθοδο που χρησιμοποιήθηκε για τον υπολογισμό της εκτίμησης της εγχώριας αγοραίας αξίας
Age-Range-18-24-F*	String		Η σημαία δηλώνει αν το νοικοκυριό περιέχει ενήλικη γυναίκα ηλικίας 18-24 ετών.
Age-Range-18-24-M*	String		Η σημαία δηλώνει αν το νοικοκυριό περιέχει ενήλικα άνδρα ηλικίας 18-24 ετών.
Age-Range-18-24-U*	String		Η σημαία υποδεικνύει εάν το νοικοκυριό περιέχει έναν ενήλικα αγνώστου φύλου ηλικίας 18-24 ετών.
Age-Range-25-34-F*	String		Η σημαία δηλώνει αν το νοικοκυριό περιέχει ενήλικη γυναίκα ηλικίας 25-34 ετών.
Age-Range-25-34-M*	String		Η σημαία δηλώνει αν το νοικοκυριό περιέχει ενήλικα άνδρα ηλικίας 25-34 ετών.
Age-Range-25-34-U*	String		Η σημαία υποδεικνύει εάν το νοικοκυριό περιέχει έναν ενήλικα αγνώστου φύλου ηλικίας 25-34 ετών.
Age-Range-35-44-F*	String		Η σημαία δηλώνει αν το νοικοκυριό περιέχει ενήλικη γυναίκα ηλικίας 35-44 ετών.
Age-Range-35-44-M*	String		Η σημαία δηλώνει αν το νοικοκυριό περιέχει ενήλικα άνδρα ηλικίας 35-44 ετών.
Age-Range-35-44-U*	String		Η σημαία υποδεικνύει εάν το νοικοκυριό περιέχει έναν ενήλικα αγνώστου φύλου ηλικίας 35-44 ετών.
Age-Range-45-54-F*	String		Η σημαία δηλώνει αν το νοικοκυριό περιέχει ενήλικη γυναίκα ηλικίας 45-54 ετών.

Age-Range-45-54-M*	String		Η σημαία δηλώνει αν το νοικοκυριό περιέχει ενήλικα άνδρα ηλικίας 45-54 ετών.
Age-Range-45-54-U*	String		Η σημαία υποδεικνύει εάν το νοικοκυριό περιέχει έναν ενήλικα αγνώστου φύλου ηλικίας 45-54 ετών.
Age-Range-55-64-F*	String		Η σημαία δηλώνει αν το νοικοκυριό περιέχει ενήλικη γυναίκα ηλικίας 55-64 ετών.
Age-Range-55-64-M*	String		Η σημαία δηλώνει αν το νοικοκυριό περιέχει ενήλικα άνδρα ηλικίας 55-64 ετών.
Age-Range-55-64-U*	String		Η σημαία υποδεικνύει εάν το νοικοκυριό περιέχει έναν ενήλικα αγνώστου φύλου ηλικίας 55-64 ετών.
Age-Range-65-74-F*	String		Η σημαία δηλώνει αν το νοικοκυριό περιέχει ενήλικη γυναίκα ηλικίας 65-74 ετών.
Age-Range-65-74-M*	String		Η σημαία δηλώνει αν το νοικοκυριό περιέχει ενήλικα άνδρα ηλικίας 65-74 ετών.
Age-Range-65-74-U*	String		Η σημαία υποδεικνύει εάν το νοικοκυριό περιέχει έναν ενήλικα αγνώστου φύλου ηλικίας 65-74 ετών.
Age-Range-over75-F*	String		Η σημαία δηλώνει αν το νοικοκυριό περιέχει ενήλικη γυναίκα ηλικίας 75 ετών και άνω.
Age-Range-over75-M*	String		Η σημαία δηλώνει αν το νοικοκυριό περιέχει ενήλικα άνδρα ηλικίας 75 ετών και άνω.
Age-Range-over75-U*	String		Η σημαία υποδεικνύει εάν το νοικοκυριό περιέχει έναν ενήλικα αγνώστου φύλου ηλικίας 75 ετών και άνω.

3.2.2 Δεδομένα Αγορών

1. Κάθε γραμμή του πίνακα αφορά ένα στοιχείο της συναλλαγής. Για να προβληθούν όλες τις αγορές που πραγματοποιήθηκαν στο πλαίσιο μιας μεμονωμένης αγοραστικής επίσκεψης ("καλάθι"), πρέπει να χρησιμοποιηθεί ο συνδυασμός των `hh_id`, `transaction-date`, `transaction-time` και `location`.
2. Το "Channel - Κανάλι" υποδεικνύει το κανάλι προέλευσης των πωλήσεων. Η εταιρεία προσφέρει μόνο πωλήσεις στο κατάστημα και online, αλλά οι online πωλήσεις κωδικοποιούνται με πληροφορίες σχετικά με την αποστολή. Οι πιθανές τιμές είναι STORE, SHPTH, BOPIS και BOSS:
 - I. STORE: Μια "τυπική" συναλλαγή λιανικής πώλησης. Ο πελάτης αγοράζει ένα προϊόν που διατίθεται στο κατάστημα σε μια φυσική τοποθεσία καταστήματος
 - II. SHPTH: Ο πελάτης αγοράζει ένα προϊόν μέσω διαδικτύου και το αποστέλλει στο σπίτι του.
 - III. BOPIS: Ο πελάτης αγοράζει ένα προϊόν online και το παραλαμβάνει σε ένα φυσικό κατάστημα
 - IV. BOSS: Ο πελάτης αγοράζει ένα προϊόν που συνήθως δεν υπάρχει στα καταστήματα και το αποστέλλει στο κατάστημα. Γενικά, υπάρχει δωρεάν αποστολή εάν το προϊόν αποστέλλεται σε μια τοποθεσία λιανικής πώλησης και όχι σε μια διεύθυνση κατοικίας, οπότε αυτή τείνει να είναι μια φθηνότερη επιλογή
 - V. Οι SHPTH, BOPIS και BOSS είναι όλες διαδικτυακές πωλήσεις και αντιπροσωπεύουν μαζί το ~,4% όλων των πωλήσεων σε αυτό το σύνολο δεδομένων.
3. Ορισμένα προϊόντα είναι διαθέσιμα μόνο διαδικτυακά, οπότε θα εμφανίζονται μόνο σε διαδικτυακά κανάλια.
4. Στο παρόν σύνολο δεδομένων περιλαμβάνονται τόσο οι αγορές όσο και οι επιστροφές. Η επιστροφή καταγράφεται ως νέα συναλλαγή, σε ξεχωριστή σειρά από την αρχική αγορά. Οι αγορές έχουν `unit_sales > 0`, οι επιστροφές έχουν `unit_sales < 0`. Η συναλλαγή που καταγράφει μια επιστροφή περιλαμβάνει επίσης την ποσότητα των επιστρεφόμενων ειδών. Το `unit_qty` δεν είναι ποτέ `< 0`. Μια επιστροφή δεν μπορεί να συνδεθεί με μια συγκεκριμένη αγορά, παρά μόνο μέσω του `item_id`. Υπάρχουν ορισμένες αγορές και επιστροφές που δεν μπορούν να αντιστοιχιστούν - ορισμένες επιστροφές αφορούν αγορές που πραγματοποιήθηκαν πριν από το παράθυρο παρατήρησης και ορισμένες αγορές μπορεί να έχουν επιστραφεί μετά το παράθυρο παρατήρησης.
5. Ενώ οι περισσότερες εκπτώσεις είναι αρνητικές τιμές (αφαιρούνται από την τελική τιμή της συναλλαγής), μια θετική τιμή σε οποιοδήποτε πεδίο έκπτωσης εμφανίζεται μόνο με επιστροφή.

6. Όλες οι εκπτώσεις εμφανίζονται σε δολάρια ΗΠΑ.
7. Όλες οι συναλλαγές καταγράφονται στην τοπική ώρα του καταστήματος/τοποθεσίας.
8. Οι πωλήσεις ανά μονάδα περιλαμβάνουν όλες τις εκπτώσεις. Είναι η καθαρή τιμή μετά την εφαρμογή όλων των εκπτώσεων
9. Οι στήλες που σημειώνονται με αστερίσκο (*) έχουν πίνακες διανομής όπου στο τέλος αυτού του πίνακα βρίσκονται μετρήσεις και συχνότητα τιμών.

Στον πίνακα 6 γίνεται παρουσίαση των δεδομένων των αγορών, μαζί με σημειώσεις για το κάθε στοιχείο.

Πίνακας 6: Δεδομένα Αγορών

Όνομα Στήλης	Τύπος Δεδομένου	Σημειώσεις SQL	Σημειώσεις / Επεξηγήσεις
Hh_id	String	Primary Key	Μοναδικό αναγνωριστικό για το νοικοκυριό
Customer-segment*	String		Ο εσωτερικός προσδιορισμός των πελατών της εταιρείας. Αυτός είναι ο μόνος τρόπος με τον οποίο η εταιρεία τμηματοποιεί τους πελάτες. 2 = επαγγελματίας, 3 = καταναλωτής (ιδιώτης)
Channel*	String		Καθορίζει την πηγή πωλήσεων, online vs. κατάστημα.
Location	String	Primary Key, Foreign Key	Το αναγνωριστικό του καταστήματος στο οποίο πραγματοποιήθηκε η συναλλαγή, παραπέμπει στον πίνακα Store
Cashier	String		Η ταμειακή μηχανή στην οποία πραγματοποιήθηκε η έγινε η συναλλαγή
Transaction-date	Date	Primary Key	Η ημερομηνία κατά την οποία πραγματοποιήθηκε η συναλλαγή
Transaction-time	Time	Primary Key	Η ώρα κατά την οποία έγινε η συναλλαγή

Item_id	String		Το αναγνωριστικό του προϊόντος, παραπέμπει στον πίνακα Product
Unit-quantity	Decimal		Η ποσότητα του εν λόγω προϊόντος που αγοράστηκε κατά τη συναλλαγή
Unit-sales	Decimal		Συνολικό ποσό που καταβλήθηκε για το στοιχείο που αγοράστηκε, συμπεριλαμβανομένης της αξίας όλων των εκπτώσεων
Loyalty-discount-1st	Decimal		Το ποσό που αφαιρείται από τη συνολική τιμή λόγω του 1ου προγράμματος πιστότητας
Loyalty-discount-2nd	Decimal		Το ποσό που αφαιρείται από τη συνολική τιμή λόγω του 2ου προγράμματος πιστότητας
Loyalty-discount-3rd	Decimal		Το ποσό που αφαιρείται από τη συνολική τιμή λόγω του 3ου προγράμματος πιστότητας
Loyalty-discount-4th	Decimal		Το ποσό που αφαιρείται από τη συνολική τιμή λόγω του 4ου προγράμματος πιστότητας
Loyalty-discount-5th	Decimal		Το ποσό που αφαιρείται από τη συνολική τιμή λόγω του 5ου προγράμματος πιστότητας
Loyalty-discount-6th	Decimal		Το ποσό που αφαιρείται από τη συνολική τιμή λόγω του 6ου προγράμματος πιστότητας
Loyalty-discount-other	Decimal		Το ποσό που αφαιρείται από τη συνολική τιμή λόγω των άλλων εκπτώσεων/προσφορών, κατά περίπτωση. Εάν το στοιχείο που αγοράστηκε ήταν σε έκπτωση, η έκπτωση θα εμφανιστεί σε αυτή τη στήλη.

Wcai_gross_sales	Decimal		Η πλήρης τιμή του στοιχείου πριν από οποιεσδήποτε μειώσεις ή εκπτώσεις
------------------	---------	--	--

Όνομα Κολώνας: cust_segment			
Τιμή	Περιγραφή	Καταμέτρηση	Ποσοστό
2	Professional Επαγγελματίας	- 9,422,747	81.02%
3	Consumer - Ιδιώτης	2,207,991	18.98%

Όνομα Κολώνας: channel		
Τιμή	Καταμέτρηση	Ποσοστό
BOPIS	9,982	0.09%
BOSS	3,569	0.03%
SHPTH	32,570	0.28%
STORE	11,584,617	99.60%

3.2.3 Δεδομένα Προϊόντων

1. Τα προσφερόμενα προϊόντα της εταιρείας είναι οργανωμένα σε μια ιεραρχία ξεκινώντας από μια ευρεία κατανομή μέχρι συγκεκριμένη:
ITEM_GRP → ITEM_CLASS → ITEM_SUBCLASS → ITEM_ID. Το ITEM_GRP είναι ο ευρύτερος τομέας του προϊόντος, π.χ. Κουζίνα, και γίνεται πιο συγκεκριμένη καθώς προχωρά προς τα κάτω στο μεμονωμένο προϊόν.
Ένα παράδειγμα της ιεραρχίας με τη σειρά που παρατίθεται παραπάνω θα μπορούσε να είναι:
Ευλεία → Σανίδες → Σκληρό Ξύλο → 2μ. Βελανιδιά
2. Ο πίνακας προϊόντων δεν περιλαμβάνει περιγραφικά χαρακτηριστικά για κάθε είδος, αλλά θα μπορούσε να γίνει εξόρυξη κειμένου για να βρεθεί μια πληθώρα πληροφοριών, συμπεριλαμβανομένων των υλικών από τα οποία αποτελείται το προϊόν (π.χ. "ξύλινη λαβή"), το μέγεθος του προϊόντος ("2x4 - 8 πόδια") ή την ποσότητα που περιλαμβάνεται σε μια συσκευασία (12 βίδες).
3. Μόνο τα προϊόντα που αγοράστηκαν τουλάχιστον μία φορά κατά τη διάρκεια του παραθύρου παρατήρησης συμπεριλήφθηκαν σε αυτόν τον κατάλογο προϊόντων. Ενδέχεται να υπάρχουν περισσότερα προϊόντα που διαθέτει η εταιρεία, προϊόντα που έχουν διακοπεί ή προϊόντα που είναι διαθέσιμα μόνο σε ορισμένα καταστήματα. Ωστόσο, δεδομένης της βαθιάς εμβέλειας των προϊόντων που παρατίθενται σε αυτό το σύνολο δεδομένων, είναι ασφαλές να γίνει υπόθεση ότι δεν υπάρχουν ανταγωνιστικά προϊόντα ή άλλα διαθέσιμα είδη.
4. Ένα κενό item_grp_cd σημαίνει ότι το προϊόν μπορεί να ανήκει στο τμήμα "Ενοικίαση εργαλείων", το οποίο δεν διαθέτει item_grp_cd.
5. Τα item_grp_cd 25, 26, 27 ή 28 μπορούν να αγνοηθούν, καθώς πρόκειται πιθανότατα για λάθος ή για καταχώρηση οικονομικής εξισορρόπησης. Τα τμήματα αυτά αποτελούν μόνο ένα μικρό μέρος των πωλήσεων.

Στον πίνακα 7 γίνεται παρουσίαση των δεδομένων των πελατών, μαζί με σημειώσεις για το κάθε στοιχείο.

Πίνακας 7: Δεδομένα Προϊόντων

Όνομα Στήλης	Τύπος Δεδομένου	Σημειώσεις SQL	Σημειώσεις / Επεξηγήσεις
Item_id	String	Primary Key	Μοναδικό αναγνωριστικό για το προϊόν
Item_desc	String		Αναλυτική περιγραφή του προϊόντος, π.χ. "2μ σανίδα Πεύκη"
Item_sc_cd	String		Κωδικός για την υποκατηγορία του προϊόντος
Item_sc_desc	String		Υποκατηγορία του προϊόντος, π.χ. "Σκληρό Ξύλο"
Item_class_cd	String		Κωδικός για την κατηγορία του στοιχείου
Item_class_desc	String		Περιγραφή της κατηγορίας του στοιχείου, π.χ. "Σανίδες"
Item_grp_cd*	String		Κωδικός για το τμήμα που το στοιχείο υπάγεται
Item_grp_desc	String		Περιγραφή του τμήματος που ανήκει το στοιχείο, π.χ. "Ξυλεία"

Όνομα Κολώνας: item_grp_cd			
Τιμή	Περιγραφή	Καταμέτρηση μοναδικών στοιχείων	Ποσοστό
028O	Κήπος/Υπαίθριο	25,888	23.13%
025H	Εξοπλισμός / Υλικό	15,220	13.60%
026P	Υδραυλικά	11,086	9.91%
027E	Ηλεκτρικά	7,660	6.84%
30	Χειροτεχνία	7,548	6.74%
028I	Κήπος/Εσωτερικού χώρου	6,273	5.61%
023F	Επένδυση τοίχου/πατώματος	6,147	5.49%
025T	Εργαλεία	6,083	5.44%
24	Χρώμα	4,975	4.45%

22	Οικοδομικά υλικά	4,775	4.27%
029K	Κουζίνα	4,543	4.06%
21	Ξυλεία	3,146	2.81%
059S	Αποθήκευση	3,004	2.68%
027L	Φωτισμός	2,704	2.42%
026B	Μπάνιο	2,654	2.37%
<i>Blank</i>	<i>Κενό</i>	195	0.17%
25	Εξοπλισμός / Υλικό	9	0.01%
28	Εποχιακά/Κήπος	6	0.01%

4 Εμπειρική Μελέτη

Στο κεφάλαιο της Εμπειρικής Μελέτης, θα παρουσιαστεί το πλαίσιο λειτουργίας της επιχείρησης και των καταναλωτών της, και στη συνέχεια, θα παρουσιαστεί διεξοδικά ο πειραματισμός της έρευνας. Ειδικότερα, θα γίνει παρουσίαση των δεδομένων των πελατών και των προϊόντων που αγόρασαν, θα παρουσιαστεί η διαδικασία της τμηματοποίησης με χρήση της RFM αλλά και της RFMV που είναι η καινούργια διάσταση, και τέλος θα γίνει παρουσίαση των αποτελεσμάτων και προσπάθεια ερμηνείας αυτών.

4.1 Πλαίσιο Λειτουργίας Επιχείρησης

Για την παρούσα έρευνα, χρησιμοποιούνται πραγματικά δεδομένα αγοραστικών καλαθιών 5.000 πελατών εταιρείας λιανικής βελτίωσης σπιτιού & ιδιοκατασκευής μαζί με όλα τα μεμονωμένα προϊόντα που αγόρασε κάθε πελάτης. Οι πελάτες επιλέχθηκαν επειδή πραγματοποίησαν τουλάχιστον μία αγορά κατά τη διάρκεια της περιόδου παρατήρησης (5 Μαρτίου 2012 - 4 Μαρτίου 2014). Επίσης, τα δεδομένα είναι από «Γνωστούς Πελάτες» όπου έγιναν "γνωστοί" όταν παρείχαν έγκυρη ταχυδρομική διεύθυνση. Εκτός από τις αγορές, το σύνολο δεδομένων περιλαμβάνει λεπτομερείς ιεραρχίες προϊόντων και χαρακτηριστικά προϊόντων, πληροφορίες για το κατάστημα και δημογραφικά για τον πελάτη. Η εταιρεία ανήκει στις 500 μεγαλύτερες εταιρείες παγκοσμίως (Fortune 500), και επιθυμεί να κρατήσει την ανωνυμία της.

Ο κλάδος της ιδιοκατασκευής – DIY ορίζεται ως οι δραστηριότητες στις οποίες τα άτομα χρησιμοποιούν ακατέργαστα και ημι-ακατέργαστα υλικά και συστατικά μέρη για να παράγουν, να μετασχηματίσουν ή να ανακατασκευάσουν υλικά αγαθά, συμπεριλαμβανομένων εκείνων που προέρχονται από το φυσικό περιβάλλον (π.χ. εξωραϊσμός). Τα καταναλωτικά προϊόντα DIY που πωλούνται από τη βιομηχανία οικιακών βελτιώσεων είναι συνήθως ημιτελή και σχεδιασμένα για χρήση σε διαφορετικές μορφές (π.χ. κομμάτια ξυλείας). Έτσι, οι άνθρωποι που αναλαμβάνουν έργα DIY (DIYers) προχωρούν πέρα από την κατασκευή του νοήματος ενός εμπορεύματος, επειδή αυτοί οι καταναλωτές είναι τόσο οι σχεδιαστές των λειτουργικών προδιαγραφών όσο και οι οικοδόμοι. Επιλέγουν μεταξύ των διαθέσιμων υλικών και εργαλείων, σχεδιάζουν τη διαδικασία εργασίας για την ολοκλήρωση του έργου και ενεργούν ως επιθεωρητές και αξιολογητές για να εξετάσουν αν έχει επιτευχθεί το τελικό αποτέλεσμα (Wolf & McQuitty, 2011)

Το 2013, η αγορά της ιδιοκατασκευής ή do-it-yourself (DIY) και των οικιακών βελτιώσεων έφτασε περίπου τα 399 δισεκατομμύρια ευρώ παγκοσμίως. Τα κίνητρα των καταναλωτών για να αναλάβουν ένα έργο DIY έχουν επίσης αλλάξει, και ενώ προηγούμενες έρευνες έχουν υποδείξει εσωτερικά κίνητρα και παράγοντες, υπάρχουν εξωτερικοί παράγοντες που επηρεάζουν τις ραγδαίες αλλαγές στην ομάδα των καταναλωτών. Η εξωτερική του οικονομικού πλούτου συνοδεύεται συχνά από μια μεγάλη τιμή και οι καταναλωτές που είναι πιο πιθανό να ασχοληθούν

με το premium DIY είναι ο ηλικιωμένος πληθυσμός που αναφέρεται ως "Baby Boomers". Αυτή η ομάδα καταναλωτών αναζητά εκλεπτυσμένο, premium τρόπο ζωής καθώς έχουν την οικονομική δυνατότητα και το κίνητρο να το κάνουν (Hatton-Jones & Teah, 2015).

Σύμφωνα με τους Wolf και McQuitty (2011) τα παραγόμενα από την αγορά κίνητρα για DIY που προτείνονται από τη μελέτη τους, είναι τα οικονομικά οφέλη, η έλλειψη ποιότητας του έτοιμου προϊόντος, η έλλειψη διαθεσιμότητας του έτοιμου προϊόντος και η ανάγκη για προσωποποίηση.

4.2 Πειραματισμός

Για την δημιουργία των μοντέλων και για την ανάλυση των δεδομένων, έχει επιλεγεί το πρόγραμμα εξόρυξης δεδομένων RapidMiner Studio.

Σε πρώτη φάση, θα δημιουργηθεί ένα μοντέλο όπου θα καταναίμει τους πελάτες σε συστάδες ανάλογα με τα χαρακτηριστικά που έχουν στις διαστάσεις RFM.

Σε δεύτερη φάση, θα κατασκευαστεί επέκταση του μοντέλου όπου θα γίνει προσπάθεια εισαγωγής και 4^{ης} διάστασης, της ποικιλίας των προϊόντων που έχουν αγοράσει οι πελάτες. Αυτή η διάσταση θα ονομαστεί "Variety" ή για συντομία "V". Το επιθυμητό αποτέλεσμα είναι η δημιουργία ενός μοντέλου RFMV όπου θα κατανέμει τους πελάτες σε συστάδες παίρνοντας υπόψη και τις 4 διαστάσεις.

4.2.1 Παρουσίαση Συνόλου Δεδομένων Πελατών

Για τις ανάγκες της ανάλυσης RFM θα χρησιμοποιηθεί το αρχείο CSV "sales" με τα δεδομένα 5.000 πελατών, μεγέθους 397.883 εγγραφών, κρατώντας τις παρακάτω 6 διαστάσεις:

1. Το μοναδικό αναγνωριστικό πελάτη (HH_ID)
2. Το μοναδικό αναγνωριστικό καταστήματος (LOCATION)
3. Η ημερομηνία επίσκεψης του πελάτη, σε μορφή HH/MM/XXXX ΩΩ:ΛΛ (DATE_TRNS)
4. Το μοναδικό αναγνωριστικό του προϊόντος που αγόρασε (ITEM_ID)
5. Η ποσότητα που αγοράστηκε από ένα συγκεκριμένο προϊόν (QUANTITY)
6. Η αξία της ποσότητας ενός συγκεκριμένου προϊόντος που αγοράστηκε σε αμερικανικά δολάρια (PRICE)

Ορισμένοι πελάτες έχουν επισκεφθεί το ίδιο ή άλλα καταστήματα περισσότερο από μια φορά την ημέρα. Πανομοιότυπες χρονικές στιγμές με το ίδιο αναγνωριστικό πελάτη στο ίδιο κατάστημα σημαίνει πως αγόρασε περισσότερα από ένα προϊόντα.

Στο δείγμα δεν υπάρχουν ελλείψεις τιμές ή χαρακτηριστικά για οποιαδήποτε διάσταση.

Η πρώτη ημερομηνία/εγγραφή είναι στις 5 Μαρτίου 2012 ώρα 7:11 το πρωί και η τελευταία είναι στις 2 Μαρτίου 2014, ώρα 7:03 το βράδυ.

Η ελάχιστη τιμή του “PRICE” ξεκινάει από \$0,01 cent του δολαρίου με μέγιστη τιμή τα \$22.978,55 δολάρια. Μέση τιμή τα \$23,66 δολάρια, με τυπική απόκλιση \$138,261.

4.2.2 Εκτέλεση Διαδικασίας RFM

Η διαδικασία που έχει ακολουθηθεί για την ομαδοποίηση των πελατών ανάλογα με τις ομοιότητές τους ως προς την επικαιρότητα, την συχνότητα και την χρηματική αξία των αγορών τους (RFM) είναι η εξής:

Ο πρώτος τελεστής (operator) είναι ο “Retrieve”, όπου ανακτά το αρχείο CSV “Πελάτες” και το φορτώνει στη διαδικασία. Σε αυτό το σύνολο δεδομένων, υπάρχουν οι πληροφορίες των πελατών για την ημερομηνία των αγορών τους, τη συνολική χρηματική αξία αυτών, το αναγνωριστικό αριθμό του κάθε πελάτη και άλλα.

Ο 2^{ος} τελεστής, ο “Multiply”, λαμβάνει τα δεδομένα από το πρώτο τελεστή στη θύρα εισόδου και παραδίδει αντίγραφα στις θύρες εξόδου. Κάθε συνδεδεμένη θύρα δημιουργεί ένα ανεξάρτητο αντίγραφο, έτσι η αλλαγή σε οποιοδήποτε από αυτά δεν επηρεάζει τα υπόλοιπα.

Τα αντίγραφα των δεδομένων έχουν μοιραστεί σε 3 τελεστές “Aggregate”, για το “M” (χρηματική αξία), για το “F” (συχνότητα αγορών) και το “R” (επικαιρότητα αγορών). Ο τελεστής “Aggregate” εκτελεί τις λειτουργίες συνάθροισης που είναι γνωστές από την SQL. Επικεντρώνεται στην απόκτηση συνοπτικών πληροφοριών όπως μέσοι όροι, αθροίσματα, μέγιστα, ελάχιστα κλπ.

Στον τελεστή “Aggregate” του M, τα δεδομένα της στήλης “Price” όπου είναι η αξία της εκάστοτε αγοράς, θα αθροιστούν για τον κάθε πελάτη. Ειδικότερα, υπάρχουν αρκετά παραδείγματα όπου πελάτες έχουν κάνει πολλαπλές αγορές την ίδια ημερομηνία.

Στον τελεστή “Aggregate” του F, οι ημερομηνίες της κάθε επίσκεψης του πελάτη θα αθροιστούν για την συχνότητα των αγορών τους.

Στον τελεστή “Aggregate” του R, η ημερομηνία επίσκεψης μετατρέπεται στη διαφορά της τελευταίας καταγεγραμμένης επίσκεψης του συνόλου, από τη τελευταία επίσκεψη του συγκεκριμένου πελάτη, μεταφρασμένη σε μήνες. Όσο μεγαλύτερες αρνητικές τιμές παίρνει το R, τόσο πιο πρόσφατη είναι η επίσκεψη του πελάτη.

Από τους τελεστές “Aggregate” του M και του F, τα δεδομένα μεταφέρονται σε 2 τελεστές “Sort M” και “Sort F” αντίστοιχα για το καθένα. Σε αυτούς τους τελεστές, τα δεδομένα που εισέρχονται ταξινομούνται σε αύξουσα ή φθίνουσα σειρά και σε αυτή τη περίπτωση, ο μοναδικός αναγνωριστικός αριθμός πελάτη ταξινομείται με αύξουσα σειρά και για το M και για το F.

Στη συνέχεια, τα τώρα ταξινομημένα δεδομένα των M και των F, περνάνε στον τελεστή “Join”, όπου αυτός ο τελεστής θα ενώσει τα δύο σύνολα χρησιμοποιώντας το αναγνωριστικό αριθμό πελάτη ως βασικό χαρακτηριστικό (key attribute).

Από τον τελεστή “Aggregate” του R, το σύνολο θα περάσει στον τελεστή “Generate Attributes” όπου η ημερομηνία των αγορών τους θα μετασχηματιστεί στη διαφορά που έχει η τελευταία τους επίσκεψη από την τελευταία ημέρα συλλογής δεδομένων της έρευνας, μεταφρασμένη σε μήνες.

Στη συνέχεια, το σύνολο του R θα μεταφερθεί στον τελεστή “Select Attributes”. Σε αυτό το σημείο, θα κρατηθεί μόνο το υποσύνολο της ημερομηνίας αγορών και τους αναγνωριστικού αριθμού πελάτη και θα διαγραφούν τα υπόλοιπα.

Το υποσύνολο θα περάσει στον τελεστή “Sort” όπου και αυτό με τη σειρά του θα ταξινομηθεί με αύξουσα σειρά με βάση το αναγνωριστικό αριθμό πελάτη.

Από τον τελεστή “Join” του M και του F, και του τελεστή “sort” του R, τα 2 σύνολα θα συνδεθούν στον τελεστή “Join MF & R” δημιουργώντας ένα καινούργιο σύνολο.

Στον επόμενη τελεστή, όπου είναι ο “Rename”, οι στήλες του αθροίσματος των επισκέψεων, του αθροίσματος της χρηματικής αξίας της κάθε επίσκεψης, και της τελευταίας επίσκεψης σε μήνες, θα μετονομαστούν σε F, M και R αντίστοιχα.

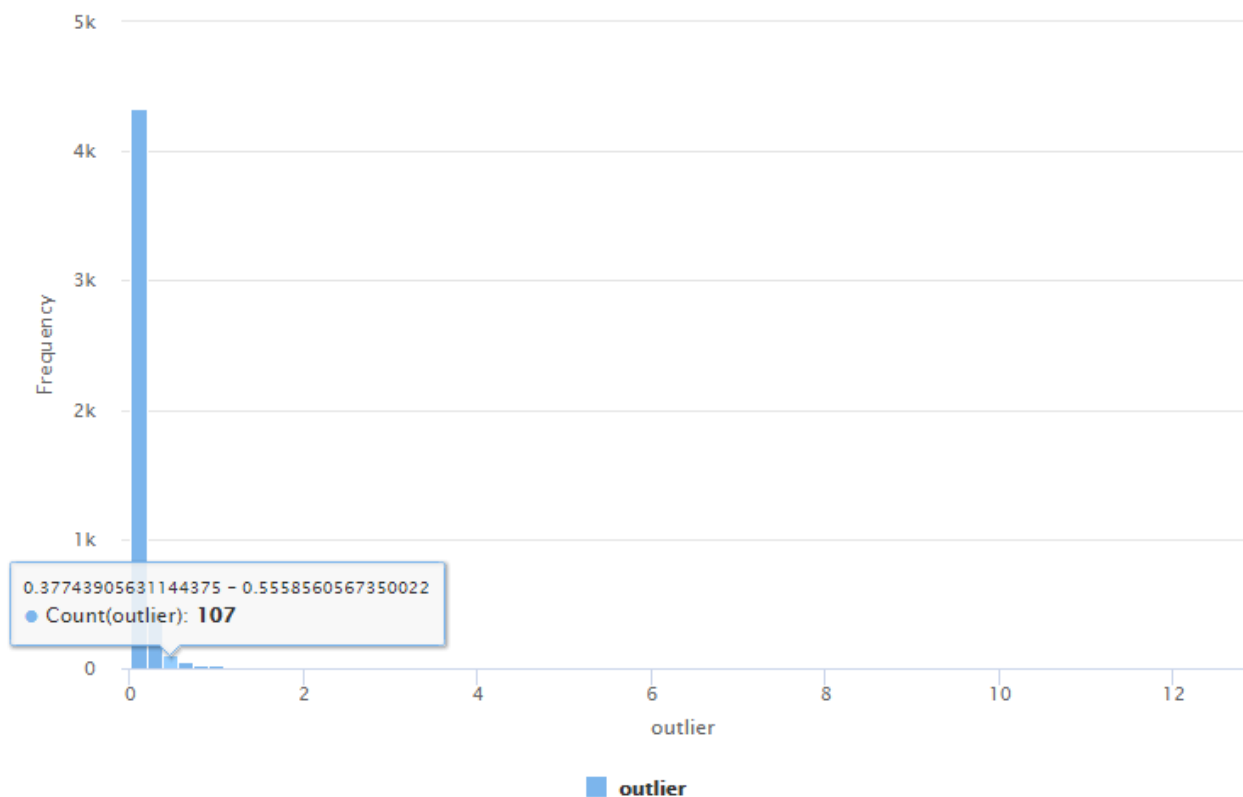
Μετά θα γίνει αναδιάταξη των στοιχείων με τη βοήθεια του τελεστή “Reorder Attributes”, σε σειρά HH_ID, R, F και M.

Από αυτά, θα δημιουργηθεί ένα υποσύνολο με τον τελεστή “Select Attributes(3)” με τα στοιχεία R, F, M.

Τα παραπάνω στοιχεία θα κανονικοποιηθούν με τον τελεστή “Normalize” καθώς οι κάθε τιμή έχει διαφορετικό πεδίο τιμών. Η μέθοδος που θα χρησιμοποιηθεί είναι η “z-Transformation”. Ονομάζεται επίσης στατιστική κανονικοποίηση. Αυτή η κανονικοποίηση αφαιρεί τη μέση τιμή των δεδομένων από όλες τις τιμές και στη συνέχεια τις διαιρεί με την τυπική απόκλιση. Στη συνέχεια, η κατανομή των δεδομένων έχει μέση τιμή μηδέν και διακύμανση ένα.

Τα κανονικοποιημένα τώρα R, F και M θα περάσουν στον τελεστή “K-NN Global Anomaly Score” όπου θα ελέγξει το σύνολο δεδομένων για ακραίες τιμές. Εφόσον όλα τα δεδομένα είναι αριθμητικά, η μεθοδολογία που θα χρησιμοποιηθεί είναι η Ευκλείδεια απόσταση, με $k=4$. Το $k=4$, που σημαίνει πως ο αριθμός γειτόνων του κάθε παραδείγματος να είναι ίσος με 4, επιλέχθηκε μετά από εμπειρική μελέτη για την επίτευξη της καλύτερη δυνατής τιμής του δείκτη Davies Bouldin και της καλύτερης δυνατής ομοιομορφίας των συστάδων.

Στη συνέχεια, θα χρησιμοποιηθεί ο τελεστής “Filter Examples” για να απομονωθούν οι ακραίες τιμές. Από αυτή την ενέργεια, διαγράφηκαν 223 πελάτες, που είχαν σκορ outlier μεγαλύτερο του 0,37743. Η πλειοψηφία του συνόλου, 4.777 πελάτες ή το 96% του δείγματος, είχαν σκορ μικρότερο του 0,37743. Το σκορ των outlier απεικονίζεται στο γράφημα 8. Οι πελάτες με ακραίες τιμές θα συγκεντρωθούν σε ξεχωριστή συστάδα ώστε να διερευνηθεί η αξία τους.



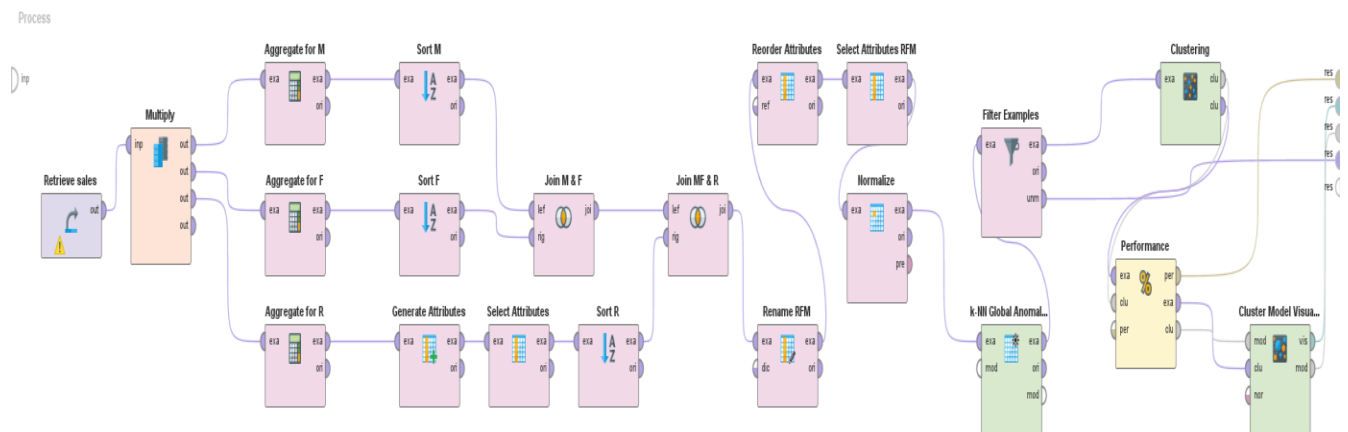
Γράφημα 8: Σκορ outlier

Αφού ξεχωριστούν οι ακραίες τιμές, το καινούργιο σύνολο είναι έτοιμο για συσταδοποίηση. Αυτό θα επιτευχθεί με τη βοήθεια του τελεστή “Clustering (k-Means)”. Αυτός ο τελεστής δημιουργεί συστάδες βάσει του αλγόριθμου k-means, όπου ελέγχει τα παραδείγματα βάσει της ομοιότητας που έχουν. Η ομοιότητα είναι αποτέλεσμα της απόστασης του ενός παραδείγματος από το άλλο. Η μεθοδολογία που ακολουθήθηκε είναι η Ευκλείδεια απόσταση, καθώς τα δεδομένα είναι αριθμητικά. Το k που επιλέχθηκε είναι το $k=3$ καθώς εξετάστηκε εμπειρικά με διάφορα k ($k=2$,

$k=3 \dots k=20$) και βρέθηκε πως το $k=3$ έχει τον καλύτερο δείκτη Davies Bouldin (-0,865). Ο αλγόριθμος k-Means θα τρέξει 30 φορές για το κάθε παράδειγμα.

Οι συστάδες που δημιουργήθηκαν θα περάσουν στο τελεστή “Performance (Cluster Distance Performance)”. Αυτός ο τελεστής χρησιμοποιείται για την αξιολόγηση των επιδόσεων των μεθόδων συσταδοποίησης με βάση το κεντροειδές. Αυτός ο τελεστής παραδίδει έναν κατάλογο τιμών κριτηρίων απόδοσης με βάση τα κεντροειδή των συστάδων. Από αυτό τον τελεστή, μεγαλύτερη σημασία έχει ο δείκτης Davies Bouldin. Η βαθμολογία Davies Bouldin ορίζεται ως το μέσο μέτρο ομοιότητας κάθε συστάδας με την πιο παρόμοια συστάδα της, όπου η ομοιότητα είναι ο λόγος των αποστάσεων εντός της συστάδας προς τις αποστάσεις μεταξύ των συστάδων. Έτσι, οι συστάδες που απέχουν περισσότερο μεταξύ τους και είναι λιγότερο διασκορπισμένες θα οδηγήσουν σε καλύτερη βαθμολογία.

Τέλος, οι συστάδες θα μοντελοποιηθούν για τη καλύτερη απεικόνιση με τη βοήθεια του τελεστή “Cluster Model Visualizer”. Αυτός ο χειριστής χρησιμοποιεί εργαλεία οπτικοποίησης για μοντέλα συστάδων με βάση το κεντροειδές για να αποτυπώσει τις βασικές λεπτομέρειες κάθε συστάδας. Οι διαδικασίες απεικονίζονται στην εικόνα 3.



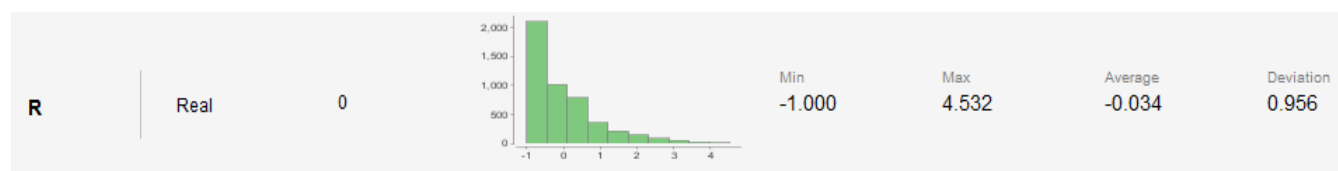
Εικόνα 3: Απεικόνιση διαδικασίας RFM στο RapidMiner Studio

4.2.3 Αποτελέσματα Διαδικασίας RFM

Με την εκτέλεση της παραπάνω διαδικασίας, το μοντέλο RFM παρουσιάζει τα παρακάτω αποτελέσματα.

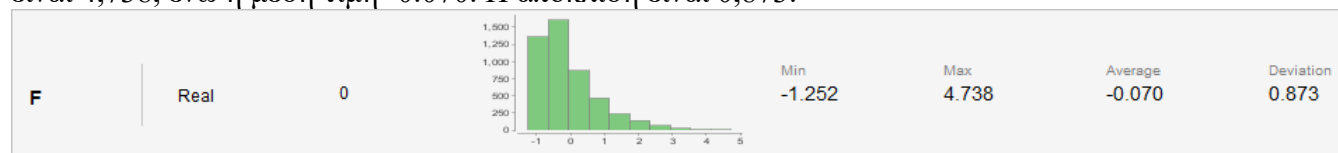
Οι κανονικοποιημένες τιμές για τις 3 διαστάσεις R, F και M παρουσιάζουν τα παρακάτω περιγραφικά στατιστικά στοιχεία, όπως φαίνεται και στις εικόνες 4, 5 και 6.

Όσον αφορά την επικαιρότητα των αγορών των πελατών (R) η ελάχιστη τιμή είναι -1.000, η μέγιστη είναι 4,532, ενώ η μέση τιμή -0.034. Η απόκλιση είναι 0,956. Οι πιο αρνητικές τιμές του R σημαίνουν ότι ο πελάτης έχει επισκεφθεί κάποιο κατάστημα πολύ πρόσφατα, ενώ οι πιο θετικές τιμές δείχνουν ότι έχει πάρα πολύ καιρό να επισκεφθεί οποιοδήποτε κατάστημα και να πραγματοποιήσει αγορά.



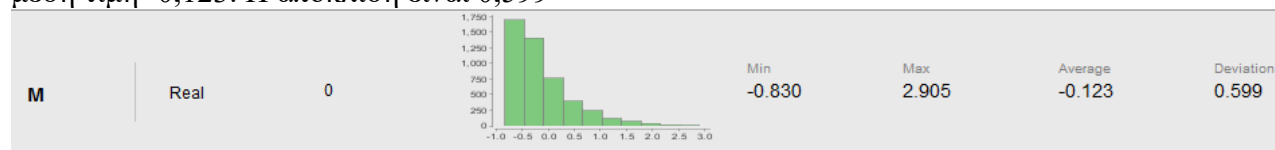
Εικόνα 4: Περιγραφικά στατιστικά για τη διάσταση R του μοντέλου RFM

Αναφορικά με την συχνότητα των επισκέψεών τους (F), η ελάχιστη τιμή είναι -1.252, η μέγιστη είναι 4,738, ενώ η μέση τιμή -0.070. Η απόκλιση είναι 0,873.



Εικόνα 5: Περιγραφικά στατιστικά για τη διάσταση F του μοντέλου RFM

Τέλος, η ελάχιστη τιμή για την χρηματική αξία των αγορών τους είναι -0,830, μέγιστη 2,905 και μέση τιμή -0,123. Η απόκλιση είναι 0,599



Εικόνα 6: Περιγραφικά στατιστικά για τη διάσταση M του μοντέλου RFM

Το μοντέλο δημιούργησε 3 συστάδες, όπως φαίνεται και στην εικόνα 7. Η πρώτη συστάδα αποτελείται από 883 πελάτες, ή δεύτερη από 1114 πελάτες, και η τρίτη και μεγαλύτερη, από 2780 πελάτες. Συνολικοί πελάτες που συσταδοποιήθηκαν είναι 4.777.

Cluster Model

```
Cluster 0: 883 items
Cluster 1: 1114 items
Cluster 2: 2780 items
Total number of items: 4777
```

Εικόνα 7: Συστάδες μοντέλου RFM

Τα κέντρα της κάθε συστάδας για τις επιμέρους διαστάσεις της R, της F και της M παρουσιάζονται στον πίνακα 8. Τα κέντρα απεικονίζουν ένα αντιπροσωπευτικό πελάτη - παράδειγμα της ανάλογης συστάδας.

Για την πρώτη συστάδα, το κέντρο της R είναι 1,601, το κέντρο της F -0.533 και το κέντρο της M -0,384.

Αντίστοιχα, για τη δεύτερη συστάδα, το κέντρο της R είναι -0,557, το κέντρο της F 1,174 και το κέντρο της M 0,619.

Τέλος, για την Τρίτη συστάδα, το κέντρο της R είναι -0,343, το κέντρο της F είναι -0,421 και το κέντρο της M -0,337.

Πίνακας 8: Κεντροειδές συστάδων για τις διαστάσεις R, F και M.

Cluster	R	F	M
Cluster 0	1.601	-0.533	-0.384
Cluster 1	-0.557	1.174	0.619
Cluster 2	-0.343	-0.421	-0.337

Για την αξιολόγηση της ομοιότητας των συστάδων, χρησιμοποιήθηκε ο δείκτης Davies Bouldin, με αποτέλεσμα -0,865. Η μέση απόσταση από το κεντροειδές είναι -0,781, ενώ ξεχωριστά για τη πρώτη συστάδα είναι -0,978, για τη δεύτερη -1,251 και για την τρίτη -0,530, όπως φαίνεται και από την εικόνα 8. Όλες οι συστάδες εμφανίζουν κοντινή απόσταση από το κέντρο, πράγμα που σημαίνει την καλή ομοιομορφία των συστάδων.

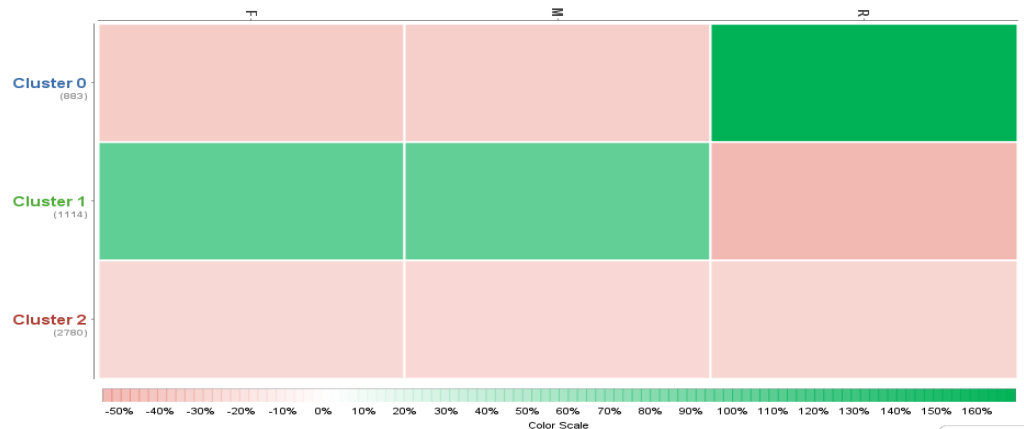
PerformanceVector

```
PerformanceVector:
Avg. within centroid distance: -0.781
Avg. within centroid distance_cluster_0: -0.978
Avg. within centroid distance_cluster_1: -1.251
Avg. within centroid distance_cluster_2: -0.530
Davies Bouldin: -0.865
```

Εικόνα 8: Αξιολόγηση ομοιότητας συστάδων με τη χρήση του δείκτη Davies Bouldin

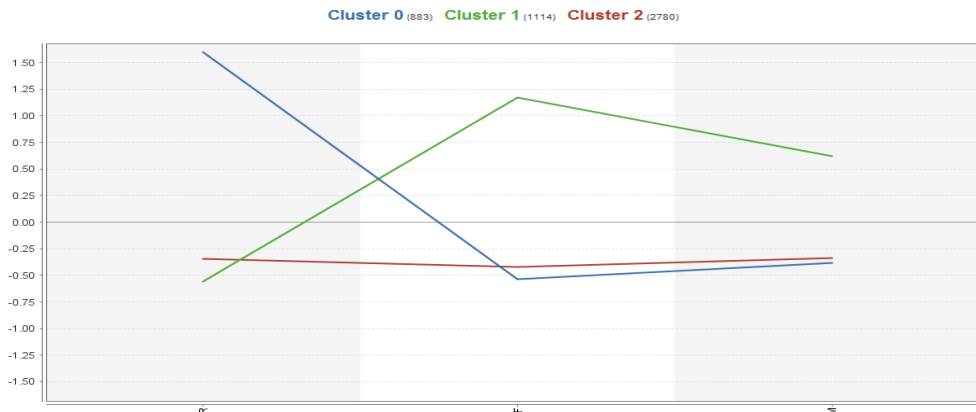
Στο χάρτη θερμότητας (γράφημα 9) απεικονίζεται η ένταση της κάθε διάστασης, για τη κάθε συστάδα.

Στη πρώτη συστάδα φαίνεται πως οι τιμές των F και των M είναι χαμηλές, αλλά οι τιμές της R αρκετά υψηλές. Για τη δεύτερη συστάδα, οι τιμές των F και των M παρουσιάζουν υψηλές τιμές, ενώ χαμηλές τιμές για την R, περισσότερο από κάθε άλλη συστάδα. Η τρίτη συστάδα παρουσιάζει χαμηλές τιμές σε όλες τις διαστάσεις.



Γράφημα 9: Χάρτης θερμότητας συστάδων για τις διαστάσεις R, F και M.

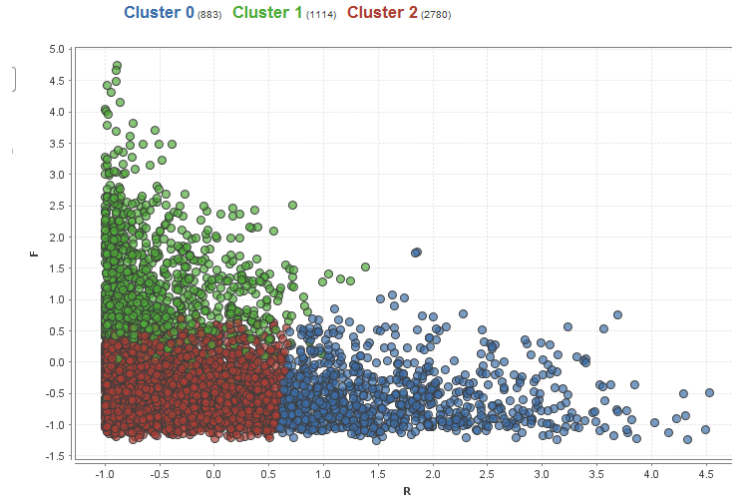
Το γράφημα 10 απεικονίζει τους αντιπροσωπευτικούς πελάτες της κάθε συστάδας. Η πρώτη συστάδα έχει τις υψηλότερες τιμές ως προς την επικαιρότητα, πράγμα που σημαίνει πως ο πελάτης δεν έχει πραγματοποιήσει πρόσφατα κάποια αγορά, και τις χαμηλότερες ως προς την συχνότητα επισκέψεων και τη χρηματική αξία των αγορών τους. Η δεύτερη συστάδα σε αντίθεση, παρουσιάζει τις χαμηλότερες τιμές ως προς την επικαιρότητα των επισκέψεων, δηλαδή ο πελάτης, έχει επισκεφθεί πρόσφατα κάποιο κατάστημα και πραγματοποιήσει αγορά, αλλά και τις υψηλότερες τιμές ως προς την συχνότητα επισκέψεων και την χρηματική αξία των αγορών τους. Η τρίτη συστάδα, παρουσιάζει χαμηλά κεντροειδές σε όλες τις διαστάσεις, και κάτω της μέσης τιμής.



Γράφημα 10: Κεντροειδές απεικόνιση διαστάσεων RFM για τις συστάδες

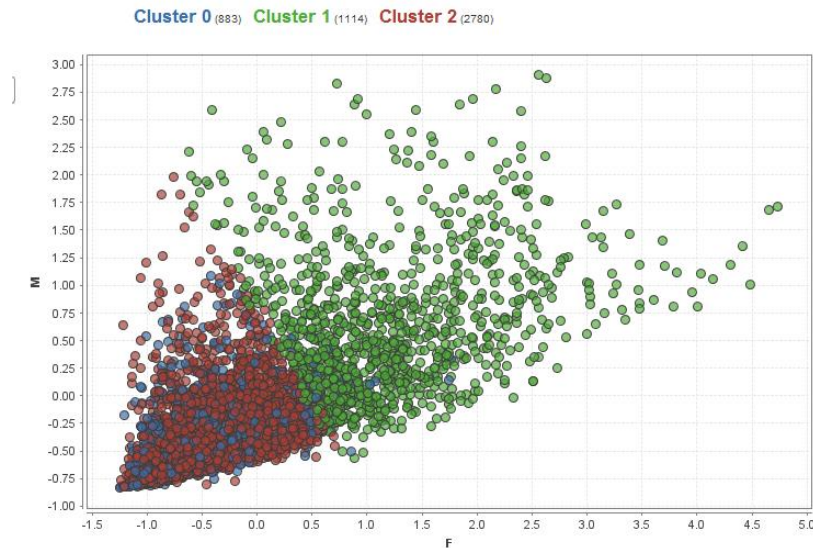
Στο παρακάτω διάγραμμα διασποράς (γράφημα 11) φαίνονται οι πυκνότητες των συστάδων ανάλογα με τις τιμές που έχουν στις διαστάσεις F και R.

Η πρώτη συστάδα εμφανίζει υψηλές τιμές για την διάσταση της R, μεγαλύτερες από αυτές των υπόλοιπων συστάδων. Η 2^η συστάδα συγκεντρώνεται σε χαμηλές τιμές για την R, αλλά με τις υψηλότερες τιμές για την F. Όσο για την τελευταία συστάδα συγκεντρώνεται σε χαμηλές τιμές, από το ελάχιστο μέχρι και την μέση τιμή τόσο για την F όσο και για την R.



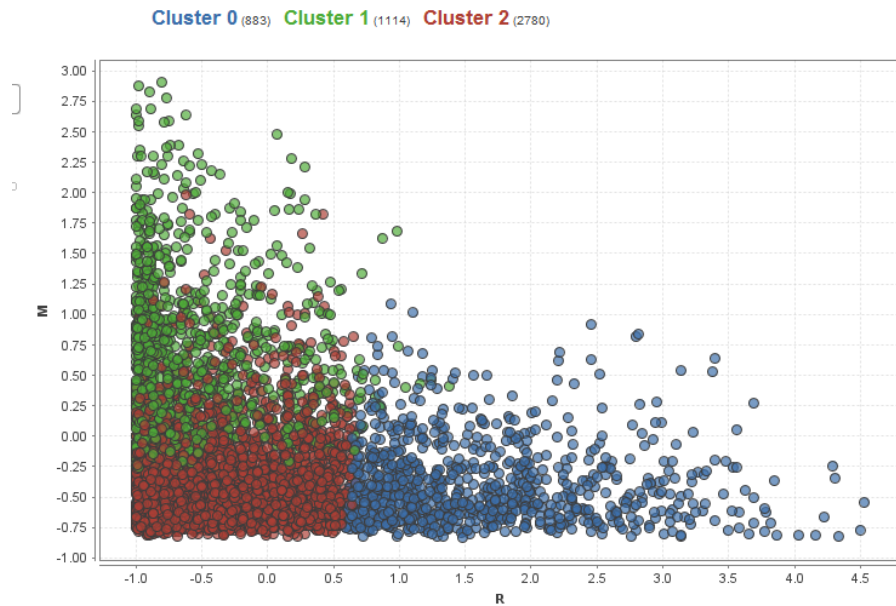
Γράφημα 11: Διάγραμμα διασποράς συστάδων ως προς το F και το R

Στο δεύτερο διάγραμμα διασποράς (γράφημα 12), παρουσιάζονται οι πυκνότητες των συστάδων ως προς τις διαστάσεις της συχνότητας επισκέψεων (F) και της χρηματικής αξίας (M). Η πρώτη και η τρίτη συστάδα εμφανίζουν έντονη πυκνότητα σε χαμηλές τιμές, τόσο για την συχνότητα όσο και για την χρηματική αξία. Η δεύτερη συστάδα ωστόσο, παρουσιάζει έντονη πυκνότητα πάνω από τις μέσες τιμές και εκτείνεται μέχρι και στα μέγιστα.



Γράφημα 12: Διάγραμμα διασποράς συστάδων ως προς το M και το F

Στο τελευταίο διάγραμμα διασποράς (γράφημα 13) παρουσιάζονται οι πυκνότητες των συστάδων ως προς τη χρηματική αξία των αγορών τους και το πόσο πρόσφατη ήταν η τελευταία τους επίσκεψη. Όπως και από το πρώτο διάγραμμα διασποράς, η πρώτη συστάδα παίρνει υψηλές τιμές ως προς την επικαιρότητα (R), και συγκεντρώνεται πάνω από τη μέση τιμή. Αντιθέτως όμως για την χρηματική αξία, όπου παίρνει χαμηλές τιμές, όπως και η τρίτη συστάδα. Η τρίτη συστάδα συγκεντρώνεται κάτω της μέσης τιμής τόσο για την χρηματική αξία (M) όσο και για την επικαιρότητα (R), περισσότερο από τις υπόλοιπες συστάδες. Η δεύτερη συστάδα είναι η μόνη συστάδα που παίρνει τις υψηλότερες τιμές για την M, και την ίδια στιγμή παρουσιάζει την μεγαλύτερη συγκέντρωση στις χαμηλότερες τιμές της R.



Γράφημα 13: Διάγραμμα διασποράς συστάδων ως προς το M και το R

Στην παρακάτω εικόνα (εικόνα 9) γίνεται παρουσίαση των συστάδων ως προς του όγκου τους και των διαστάσεων τους RFM.

Η πρώτη και μικρότερη συστάδα, όγκου 883 πελατών έχει κατά μέσο όρο 169% μεγαλύτερες τιμές R από τις υπόλοιπες συστάδες, συχνότητα επισκέψεων κατά μέσο όρο 39% μικρότερη, και χρηματική αξία μικρότερη κατά μέσο όρο κοντά στα 37% από τις άλλες 2 συστάδες.

Η δεύτερη συστάδα, όγκου 1.114 πελατών, είναι η συστάδα με τις πιο πρόσφατες επισκέψεις, κατά μέσο όρο 54% μικρότερο R, την μεγαλύτερη συχνότητα επισκέψεων με μέσο όρο 105%, και τη μεγαλύτερη χρηματική αξία αγορών με μέσο όρο 104% μεγαλύτερο από τις υπόλοιπες συστάδες.

Η τρίτη συστάδα αποτελείται με πάνω από τους μισούς πελάτες του δείγματος, όγκου 2,780 πελατών. Σε αυτή τη συστάδα βρίσκονται πελάτες που έχουν πραγματοποιήσει σχετικά πρόσφατα αγορά, με μέσο όρο 32% μικρότερο R από τις υπόλοιπες συστάδες. Η χρηματική αξία και η συχνότητα των αγορών τους είναι κατά μέσο όρο περίπου 30% μικρότερες από τις άλλες 2 συστάδες.

Number of Clusters: 3

Cluster 0

883

R is on average 169.09% larger, F is on average 39.21% smaller, M is on average 36.98% smaller

Cluster 1

1,114

F is on average 105.22% larger, M is on average 104.93% larger, R is on average 54.18% smaller

Cluster 2

2,780

R is on average 32.00% smaller, M is on average 30.30% smaller, F is on average 29.71% smaller

Εικόνα 9: Συστάδες RFM

4.2.4 Ερμηνεία Αποτελεσμάτων RFM

Από τα αποτελέσματα του μοντέλου της RFM, δημιουργήθηκαν 3 συστάδες.

Είναι εμφανής η συσχέτιση μεταξύ της συνολικής χρηματικής αξίας που ξόδεψαν οι πελάτες και της συχνότητας των αγορών.

Η πρώτη συστάδα (cluster 0) αποτελείται από πελάτες οι οποίοι δεν έχουν επισκεφθεί κάποιο κατάστημα για αγορά, για πάρα πολύ καιρό. Επίσης, οι συχνότητα των αγορών τους αλλά και η χρηματική αξία αυτών είναι η χαμηλότερη σε σύγκριση με τις υπόλοιπες δύο συστάδες. Θεωρείται πως οι συγκεκριμένοι πελάτες είχαν αναλάβει ένα συγκεκριμένο έργο ιδιοκατασκευής το οποίο έχει πλέον ολοκληρωθεί, οπότε σταμάτησαν να επισκέπτονται και το κατάστημα για την αγορά εξοπλισμού και υλικών. Επίσης, εάν υπολογιστεί και η επιρροή του ανταγωνισμού, τότε αυτή η συστάδα πελατών, η οποία δεν έχει πραγματοποιήσει πρόσφατα κάποια αγορά, μπορεί να συνεχίζει τις αγορές της σε άλλη παρόμοια εταιρεία, ή τέλος, να έκανε κάποιες ομορτυνιστικές αγορές στην εταιρεία της έρευνας λόγω κάποιας ευνοϊκής για αυτούς προσφορά.

Η δεύτερη συστάδα (cluster 1), εμφανίζει σε πρώτη ματιά τον ιδανικό πελάτη ως προς τις διαστάσεις RFM. Η συγκεκριμένη συστάδα αποτελείται από πελάτες οι οποίοι κάνουν τις συχνότερες αγορές, ξοδεύουν τα περισσότερα χρήματα και η τελευταία τους επίσκεψη είναι παρά πολύ πρόσφατη. Παίρνοντας υπόψη ότι η περίοδος παρατήρησης είναι 2 χρόνια, και οι συγκεκριμένοι πελάτες είναι ενεργοί καθ' όλη τη διάρκεια, μπορεί να ειπωθεί ότι αυτή η συστάδα δείχνει μερική πιστότητα προς την επιχείρηση. Ενέργειες θα πρέπει να γίνουν για τη διατήρηση αυτής.

Η τρίτη και μεγαλύτερη συστάδα είναι ίσως η πιο σημαντική και η πιο πολύτιμη για την εταιρεία. Οι πελάτες αυτής της συστάδας εμφανίζουν σχετικά χαμηλή συχνότητα επισκέψεων και η χρηματική αξία των αγορών τους είναι επίσης χαμηλή συγκριτικά με την 2^η συστάδα. Όμως, η τελευταία τους αγορά είναι αρκετά πρόσφατη. Η συγκεκριμένη συστάδα που αποτελείται από πάνω από τους μισούς πελάτες του δείγματος, οι οποίοι δείχνουν μια καλή συχνότητα αγορών και σχετικά καλή χρηματική αξία αυτών, φέρουν τη μεγαλύτερη και την πιο σταθερή χρηματοροή στην εταιρεία. Είναι σημαντικό να αναφερθεί επίσης ότι η συγκεκριμένη συστάδα έχει πολύ καλή ομοιομορφία των χαρακτηριστικών των πελατών όπου αποτελείται, καθώς η μέση απόσταση από το κέντρο είναι η χαμηλότερη (-0,530).

4.2.5 Παρουσίαση Συνόλου Δεδομένων Προϊόντων

Για τις ανάγκες της ανάλυσης RFMV θα χρησιμοποιηθεί το αρχείο CSV “products” με τα δεδομένα των 5.000 πελατών, μεγέθους 111.916 εγγραφών, κρατώντας τις παρακάτω 6 διαστάσεις:

1. Τον αναγνωριστικό αριθμό προϊόντος (ITEM_ID)
2. Την περιγραφή του αναγνωριστικού αριθμού προϊόντος (ITEM_DESC)
3. Τον αναγνωριστικό αριθμό υποκατηγορίας του προϊόντος (ITEM_SC_ID)
4. Την περιγραφή του αναγνωριστικού αριθμού υποκατηγορίας προϊόντος (ITEM_SC_DESC)
5. Τον αναγνωριστικό αριθμό κατηγορίας του προϊόντος (ITEM_CLASS_CD)
6. Την περιγραφή του αναγνωριστικού αριθμού κατηγορίας προϊόντος (ITEM_CLASS_DESC)
7. Τον αναγνωριστικό κωδικό τομέα του προϊόντος (ITEM_GRP_CD)
8. Την περιγραφή του αναγνωριστικού αριθμού τομέα του προϊόντος (ITEM_GRP_DESC)

Τα προσφερόμενα προϊόντα της εταιρείας είναι οργανωμένα σε μια ιεραρχία ξεκινώντας από μια ευρεία κατανομή μέχρι συγκεκριμένη:

ITEM_GRP → ITEM_CLASS → ITEM_SUBCLASS → ITEM_ID. Το ITEM_GRP είναι ο ευρύτερος τομέας του προϊόντος, π.χ. Κουζίνα, και γίνεται πιο συγκεκριμένη καθώς προχωρά προς τα κάτω στο μεμονωμένο προϊόν.

Ένα παράδειγμα της ιεραρχίας με τη σειρά που παρατίθεται παραπάνω θα μπορούσε να είναι: Ξυλεία → Σανίδες → Σκληρό Ξύλο → 2μ. Βελανιδιά.

Οι περιγραφές είναι η ονομασία του προϊόντος ή της κατηγορίας που ανήκει το προϊόν

4.2.6 Εκτέλεση Διαδικασίας RFMV

Η διαδικασία που έχει ακολουθηθεί για την RFMV είναι επέκταση της RFM. Ακολουθούν τα βήματα.

Αρχικά, γίνεται ανάκτηση του αρχείου “products” με τον τελεστή “Retrieve”.

Το αρχείο στη συνέχεια θα περάσει στον τελεστή “Numerical to Polynominal”. Σε αυτό τον τελεστή, ένα υποσύνολο τιμών, συγκεκριμένα οι τιμές του “ITEM_CLASS_CD”, του “ITEM_SC_CD” και του “ITEM_ID”, από αριθμητικές θα μετατραπούν σε πολυωνυμικές.

Το αρχείο σε επόμενο στάδιο θα περάσει στον τελεστή “Generate Attributes (Product Category)”. Με τη βοήθεια αυτού του τελεστή, οι κωδικοί του τομέα του προϊόντος (ITEM_GRP_CD), της κατηγορίας (ITEM_CLASS_CD) και της υποκατηγορίας (ITEM_SC_CD) θα συνδεθούν και θα δημιουργηθεί ένας καινούργιος κωδικός, με την ονομασία “Product Category”. Όλα τα υπόλοιπα

χαρακτηριστικά του συνόλου θα κρατηθούν ως έχουν στο αρχείο, και θα προστεθεί ο νέος κωδικός.

Την ίδια στιγμή, από το αρχείο του “sales” και συγκεκριμένα από τον τελεστή “Multiply” του 2^{ου} βήματος της αρχικής διαδικασίας RFM, θα παραδώσει ένα ακόμα αντίγραφο, εκτός από τα 3 για τις ανάγκες τις RFM, όπου θα περάσει στον τελεστή “Numerical to Polynomial ITEM_ID”. Από το αρχείο “sales”, οι τιμές του “ITEM_ID” θα μετατραπούν από αριθμητικές σε πολυωνυμικές, και μαζί με τα υπόλοιπα στοιχεία, όπως το “HH_ID” του πελάτη και η ημερομηνία αγοράς “DATE_TRNS, θα περάσουν στον τελεστή “Join(4)” της επέκτασης RFMV.

Οι έξοδοι από τους 2 τελεστές θα περάσουν στον τελεστή “Join Sales & Products” όπου μέσω μια εσωτερικής σύνδεσης των “ITEM_ID” από το αρχείο “sales” και του “ITEM_ID” από το αρχείο “products” σαν πρωτεύων κλειδιά, θα ενωθούν ώστε να γίνει συσχετισμός των πελατών, με τα προϊόντα που έχουν αγοράσει.

Το ενωμένο αρχείο, θα περάσει στον τελεστή “Select Attributes (V, Date, ID)” όπου θα δημιουργήσει ένα υποσύνολο με τα στοιχεία “DATE_TRNS” για την ημερομηνία αγοράς, το “ID” του πελάτη και το “Product_Category”, το νέο χαρακτηριστικό που δημιουργήθηκε και περιέχει σε ένα χαρακτηριστικό τον τομέα, την κατηγορία και την υποκατηγορία του προϊόντος που αγόρασε ο πελάτης.

Στη συνέχεια, το νέο υποσύνολο θα περάσει στον τελεστή “Aggregate for V”. Εδώ, οι τιμές του χαρακτηριστικού “Product_Category” θα αθροιστούν, θα κρατηθούν οι μοναδικές τιμές για τον κάθε πελάτη και θα ομαδοποιηθούν βάση του χαρακτηριστικού “ID”, τον μοναδικό αναγνωριστικό αριθμό πελάτη. Αυτό το βήμα είναι το πιο κύριο για την επέκταση του μοντέλου, καθώς θα εμφανίσει την ποικιλία προϊόντων που έχει αγοράσει ο κάθε πελάτης.

Στο επόμενο βήμα, η επέκταση του μοντέλου, και συγκεκριμένα το κομμάτι της “V”, θα ενωθεί με το υπόλοιπο μοντέλο της RFM, καθώς έχει κρατηθεί ως έχει η διαδικασία μέχρι και αυτό το σημείο. Ο τελεστής “Join RFM & V” θα πάρει το ενωμένο σύνολο RFM από τον τελεστή “Join MF & R” και θα το συνδέσει με την έξοδο του τελεστή “Aggregate for V” μέσω εσωτερικής σύνδεσης και θα διαγράψει διπλά χαρακτηριστικά που θα υπάρξουν.

Τα χαρακτηριστικά του αθροίσματος των επισκέψεων, του αθροίσματος της χρηματικής αξίας της κάθε επίσκεψης, της τελευταίας επίσκεψης του πελάτη σε μήνες και η ποικιλία προϊόντων που αγόρασε ο πελάτης, θα μετονομαστούν σε R, F, M και V αντίστοιχα, με τη βοήθεια του τελεστή “Rename”.

Συγκεκριμένα:

1. $\text{Count}(\text{DATE_TRNS}) = F$
2. $\text{Sum}(\text{PRICE}) = M$
3. $R_IN_MONTHS = R$
4. $\text{Count}(\text{Product_Category}) = V$

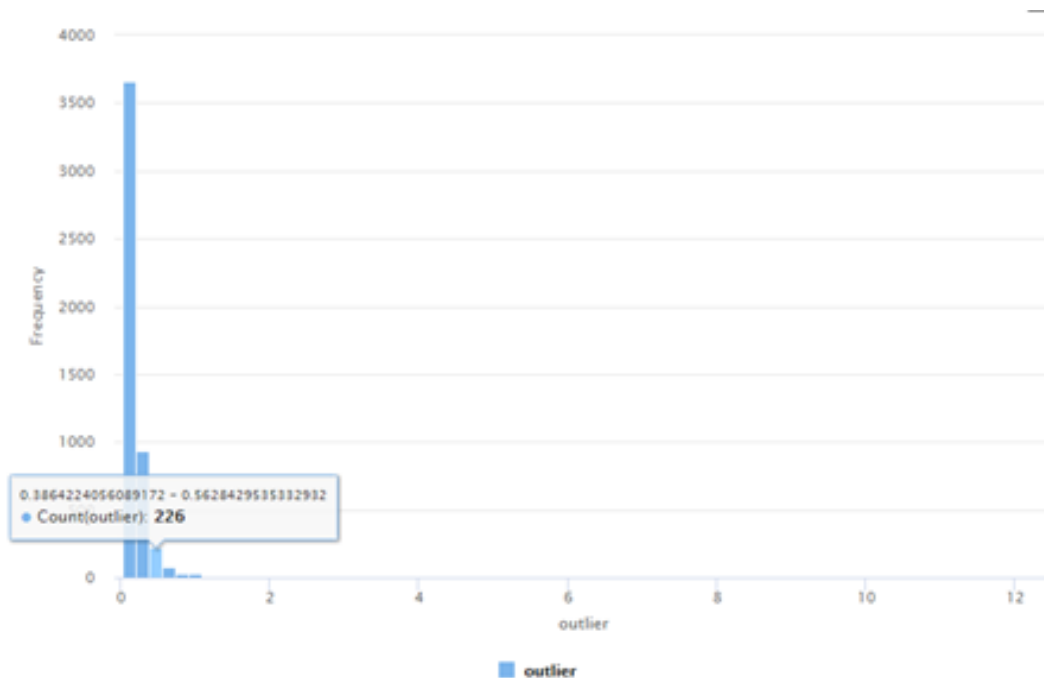
Στη συνέχεια θα γίνει αναδιάταξη των τιμών των χαρακτηριστικών “ID”, R, F, M και V, με αύξουσα σειρά, με τη βοήθεια του τελεστή “Reorder Attributes”.

Από αυτό το σύνολο, θα δημιουργηθεί υποσύνολο χρησιμοποιώντας τον τελεστή “Select Attributes (RFMV)”. Θα κρατηθούν οι τιμές R, F, M και V εκτός του “HH_ID”.

Το υποσύνολο θα κανονικοποιηθεί με τη χρήση του τελεστή “Normalize” καθώς οι τιμές περιέχουν ημερομηνίες σε μήνες, χρηματική αξία σε δολάρια και άλλα. Η μέθοδος που θα χρησιμοποιηθεί για την κανονικοποίηση είναι η z-transformation, ή αλλιώς στατιστική κανονικοποίηση. Αυτή η κανονικοποίηση αφαιρεί τη μέση τιμή των δεδομένων από όλες τις τιμές και στη συνέχεια τις διαιρεί με την τυπική απόκλιση. Στη συνέχεια, η κατανομή των δεδομένων έχει μέση τιμή μηδέν και διακύμανση ένα.

Τα πλέον κανονικοποιημένα δεδομένα των χαρακτηριστικών θα ελεγχθούν για ακραίες τιμές με τη χρήση του τελεστή “k-NN Global Anomaly Score”. Το k που έχει επιλεγεί είναι 3, δηλαδή ο αριθμός γειτόνων που παίρνονται υπόψη για το κάθε παράδειγμα, όπου μετά από εμπειρική εξέταση είχε τα καλύτερα αποτελέσματα όσον αφορά τον δείκτη Davies Bouldin που θα παρουσιαστεί στη συνέχεια. Εφόσον τα δεδομένα είναι αριθμητικής φύσεως, θα χρησιμοποιηθεί η Ευκλείδεια απόσταση για την εύρεση των ακραίων τιμών. Εφόσον όλα τα δεδομένα είναι αριθμητικά, η μεθοδολογία που θα χρησιμοποιηθεί είναι η Ευκλείδεια απόσταση, με $k=4$. Το $k=4$, που σημαίνει πως ο αριθμός γειτόνων του κάθε πελάτη να είναι ίσος με 4, επιλέχθηκε μετά από εμπειρική μελέτη για την επίτευξη της καλύτερης δυνατής τιμής του δείκτη Davies Bouldin και της καλύτερης δυνατής ομοιομορφίας των συστάδων.

Στο σύνολο θα χρησιμοποιηθεί ο τελεστής “Filter Examples” για να απομονωθούν οι ακραίες τιμές. Από αυτή την ενέργεια, διαγράφηκαν 228 πελάτες, που είχαν σκορ outlier μεγαλύτερο του 0,53. Η πλειοψηφία του συνόλου, 4.772 πελάτες ή το 96% του δείγματος, είχαν σκορ μικρότερο του 0,56. Το σκορ των outlier απεικονίζεται στο γράφημα 14. Οι πελάτες με ακραίες τιμές θα συγκεντρωθούν σε ξεχωριστή συστάδα ώστε να διερευνηθεί η αξία τους.

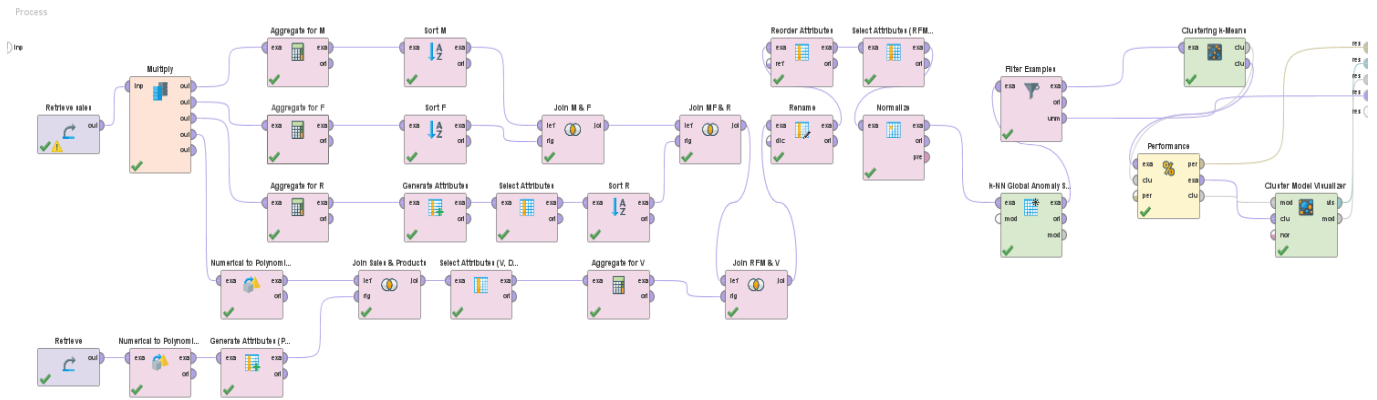


Γράφημα 14: Σκορ outlier της διαδικασίας RFMV

Το φιλτραρισμένο και κανονικοποιημένο πλέον σύνολο είναι έτοιμο για συσταδοποίηση. Αυτό θα επιτευχθεί με τη χρήση του τελεστή “Clustering k-Means”. Αυτός ο τελεστής δημιουργεί συστάδες βάσει του αλγόριθμου k-means, όπου ελέγχει τα παραδείγματα βάσει της ομοιότητας που έχουν. Η ομοιότητα είναι αποτέλεσμα της απόστασης του ενός παραδείγματος από το άλλο. Η μεθοδολογία που ακολουθήθηκε είναι η Ευκλείδεια απόσταση, καθώς τα δεδομένα είναι αριθμητικά. Το k που επιλέχθηκε είναι το k=3 καθώς εξετάστηκε εμπειρικά με διάφορα k (k=2, k=3...k=20) και βρέθηκε πως το k=3 έχει τον καλύτερο δείκτη Davies Bouldin (-0,889). Ο αλγόριθμος k-Means θα τρέξει 30 φορές για το κάθε παράδειγμα.

Οι συστάδες που δημιουργήθηκαν θα περάσουν στο τελεστή “Performance (Cluster Distance Performance)”. Αυτός ο τελεστής χρησιμοποιείται για την αξιολόγηση των επιδόσεων των μεθόδων συσταδοποίησης με βάση το κεντροειδές. Αυτός ο τελεστής παραδίδει έναν κατάλογο τιμών κριτηρίων απόδοσης με βάση τα κεντροειδή των συστάδων, με πιο σημαντικό δείκτη των Davies Bouldin. Η βαθμολογία Davies Bouldin ορίζεται ως το μέσο μέτρο ομοιότητας κάθε συστάδας με την πιο παρόμοια συστάδα της, όπου η ομοιότητα είναι ο λόγος των αποστάσεων εντός της συστάδας προς τις αποστάσεις μεταξύ των συστάδων. Έτσι, οι συστάδες που απέχουν περισσότερο μεταξύ τους και είναι λιγότερο διασκορπισμένες θα οδηγήσουν σε καλύτερη βαθμολογία.

Τέλος, οι συστάδες θα μοντελοποιηθούν για τη καλύτερη απεικόνιση με τη βοήθεια του τελεστή “Cluster Model Visualizer”. Αυτός ο χειριστής χρησιμοποιεί εργαλεία οπτικοποίησης για μοντέλα συστάδων με βάση το κεντροειδές για να αποτυπώσει τις βασικές λεπτομέρειες κάθε συστάδας. Οι διαδικασία απεικονίζεται στην εικόνα 10.



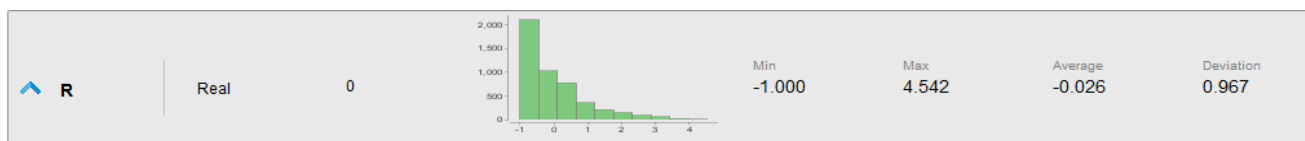
Εικόνα 10: Απεικόνιση διαδικασίας RFMV με τη χρήση του RapidMiner Studio

4.2.7 Αποτελέσματα Διαδικασίας RFMV

Με την εκτέλεση της παραπάνω διαδικασίας RFMV, το μοντέλο παρουσιάζει τα παρακάτω αποτελέσματα.

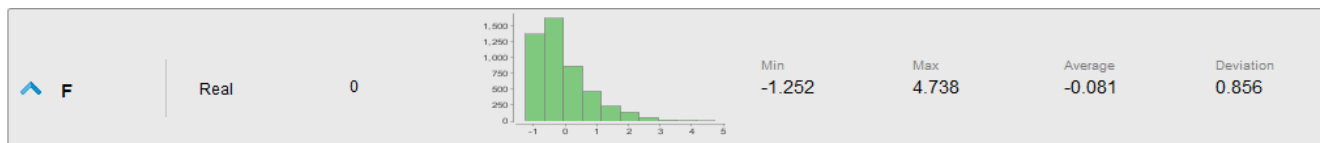
Οι κανονικοποιημένες τιμές για τις 4 διαστάσεις R, F, M και V παρουσιάζουν τα παρακάτω περιγραφικά στατιστικά στοιχεία, όπως φαίνεται και στις εικόνες 11, 12, 13 και 14.

Όσον αφορά την επικαιρότητα των αγορών των πελατών (R) η ελάχιστη τιμή είναι -1.000, η μέγιστη είναι 4,542, ενώ η μέση τιμή -0.026. Όπως και από το αρχικό μοντέλο RFM, οι πιο αρνητικές τιμές του R σημαίνουν ότι ο πελάτης έχει επισκεφθεί κάποιο κατάστημα πολύ πρόσφατα, ενώ οι πιο θετικές τιμές δείχνουν ότι έχει πάρα πολύ καιρό να επισκεφθεί οποιοδήποτε κατάστημα.



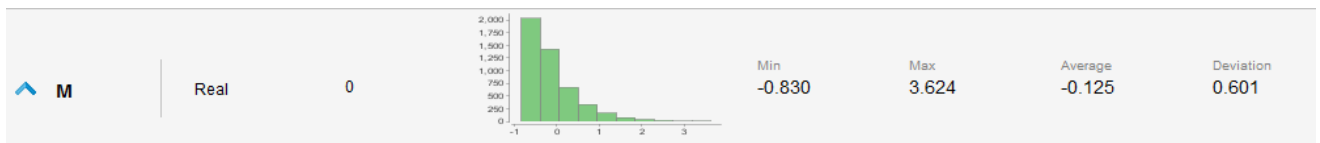
Εικόνα 11: Περιγραφικά στοιχεία για τη διάσταση R του μοντέλου RFMV

Αναφορικά με την συχνότητα των επισκέψεων τους (F), η ελάχιστη τιμή είναι -1.252, η μέγιστη είναι 4,738, ενώ η μέση τιμή -0.081. Η απόκλιση είναι 0,856.



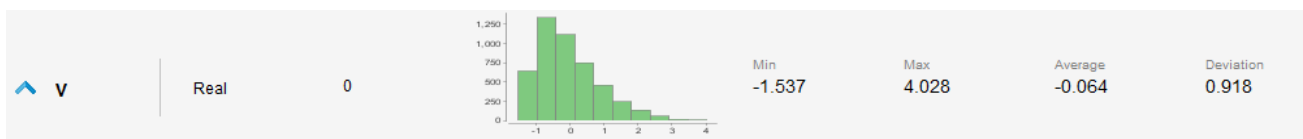
Εικόνα 12: Περιγραφικά στοιχεία για τη διάσταση F του μοντέλου RFMV

Για την διάσταση M, η ελάχιστη τιμή για την χρηματική αξία των αγορών τους είναι -0,830, μέγιστη 3,624 και μέση τιμή -0,125. Η απόκλιση είναι 0,601. Σημαντικό να αναφερθεί είναι πως η μέγιστη χρηματική αξία των αγορών αυξήθηκε κατά σχεδόν 1 μονάδα σε σχέση με το αρχικό μοντέλο RFM.



Εικόνα 13: Περιγραφικά στοιχεία για τη διάσταση M του μοντέλου RFMV

Τέλος, η επέκταση του μοντέλου, η διάσταση V, η ελάχιστη τιμή για την ποικιλία προϊόντων που αγόρασαν οι πελάτες είναι -1,537, μέγιστη είναι 4.028 και μέση τιμή -0.064. Η απόκλιση είναι 0,918.



Εικόνα 14: Περιγραφικά στοιχεία για τη διάσταση V του μοντέλου RFMV

Το καινούργιο μοντέλο δημιούργησε και αυτό 3 συστάδες, όπως φαίνεται και στην εικόνα 15. Η πρώτη συστάδα (Cluster_0) αποτελείται από 1230 πελάτες, ή δεύτερη (Cluster_1) από 847 πελάτες, και η τρίτη και μεγαλύτερη (Cluster_2), από 2695 πελάτες. Συνολικοί πελάτες που συσταδοποιήθηκαν είναι 4.772.

Cluster Model

```
Cluster 0: 1230 items
Cluster 1: 847 items
Cluster 2: 2695 items
Total number of items: 4772
```

Εικόνα 15: Συστάδες μοντέλου RFMV

Τα κέντρα της κάθε συστάδας για τις επιμέρους διαστάσεις της R, της F, της M και της V, παρουσιάζονται στον πίνακα 9. Τα κέντρα απεικονίζουν ένα αντιπροσωπευτικό πελάτη - παράδειγμα της ανάλογης συστάδας.

Για την πρώτη συστάδα, το κέντρο της R είναι -0,511, το κέντρο της F 1,078, το κέντρο της M 0,500 και για την καινούργια διάσταση, το κέντρο της V είναι 1,176.

Αντίστοιχα, για τη δεύτερη συστάδα, το κέντρο της R είναι 1,658, το κέντρο της F είναι -0,561, το κέντρο της M είναι -0,393 και της V είναι -0,614.

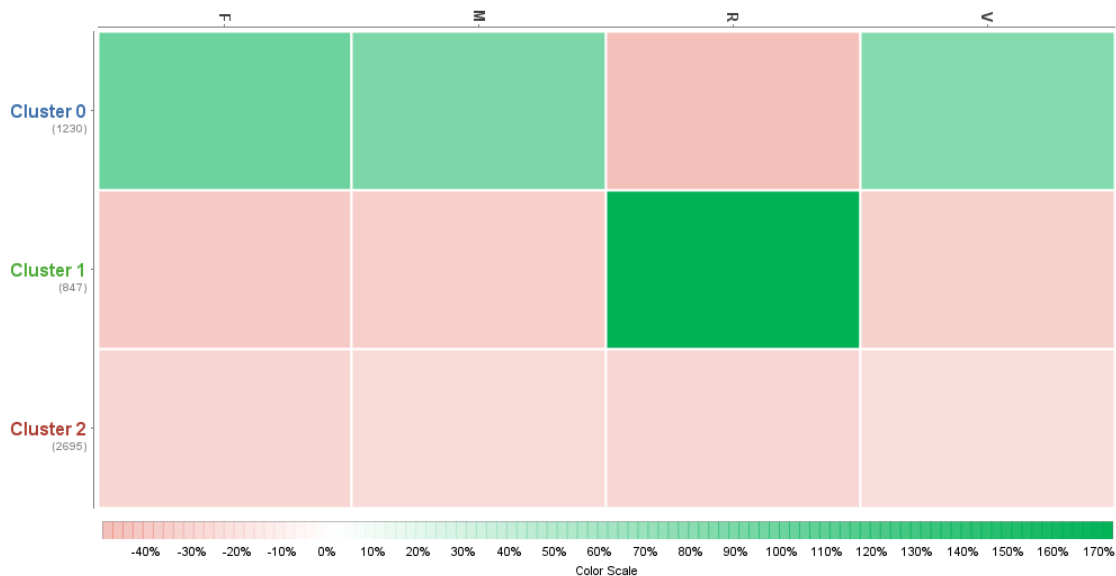
Τέλος, για την τρίτη συστάδα, το κέντρο της R είναι -0,334, το κέντρο της F είναι -0,459, το κέντρο της M -0,325 και της V -0,457.

Πίνακας 9: Κεντροειδές συστάδων για τις διαστάσεις R, F, M, και V

Cluster	R	F	M	V
Cluster 0	-0.511	1.078	0.500	1.176
Cluster 1	1.658	-0.561	-0.393	-0.614
Cluster 2	-0.334	-0.459	-0.325	-0.457

Στο χάρτη θερμότητας (γράφημα 15) απεικονίζεται η ένταση της κάθε διάστασης, για τη κάθε συστάδα.

Στη πρώτη συστάδα φαίνεται πως οι τιμές των F, των M και την V είναι υψηλές, και την ίδια στιγμή, οι τιμές της R αρκετά χαμηλές. Για τη δεύτερη συστάδα, οι τιμές των F, των M και την V παρουσιάζουν χαμηλές τιμές, ενώ υψηλές τιμές για την R, περισσότερο από κάθε άλλη συστάδα. Η τρίτη συστάδα παρουσιάζει χαμηλές τιμές σε όλες τις διαστάσεις.



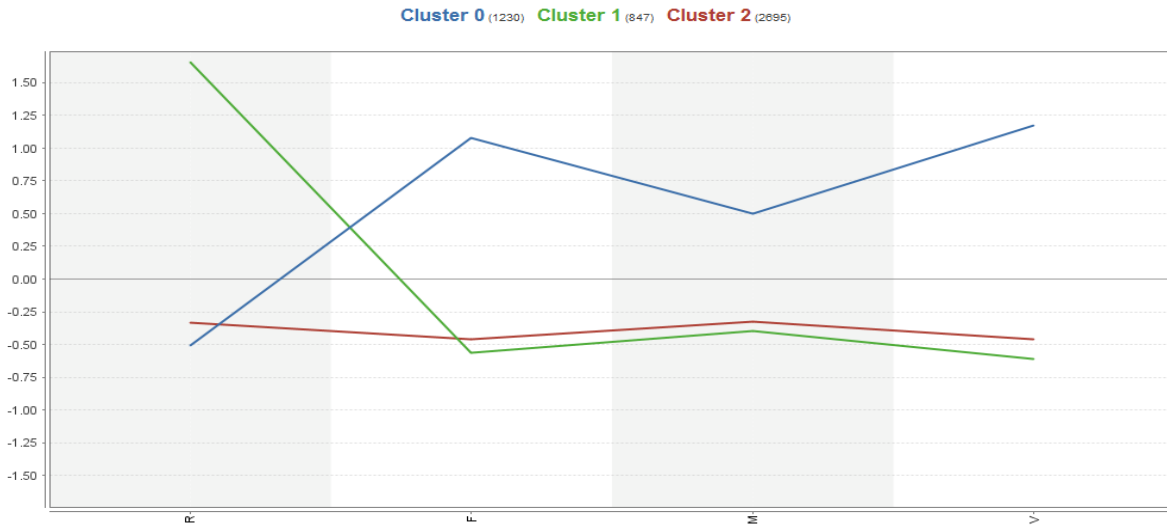
Γράφημα 15: Χάρτης θερμότητας συστάδων για τις διαστάσεις R, F, M και V

Το γράφημα 16 απεικονίζει τους αντιπροσωπευτικούς πελάτες της κάθε συστάδας.

Η πρώτη συστάδα έχει τις χαμηλότερες τιμές ως προς την επικαιρότητα, πράγμα που σημαίνει πως ο πελάτης έχει πραγματοποιήσει πρόσφατα κάποια αγορά, και τις υψηλότερες ως προς την συχνότητα επισκέψεων, τη χρηματική αξία των αγορών τους και την ποικιλία προϊόντων που έχουν αγοράσει.

Η δεύτερη συστάδα σε αντίθεση, παρουσιάζει τις υψηλότερες τιμές ως προς την επικαιρότητα των επισκέψεων, δηλαδή ο πελάτης, δεν έχει επισκεφθεί πρόσφατα κάποιο κατάστημα και δεν πραγματοποίησε αγορά, και τις χαμηλότερες τιμές τόσο για την συχνότητα επισκέψεων, όσο και για την χρηματική αξία των αγορών τους αλλά και της ποικιλίας των προϊόντων που έχουν αγοράσει.

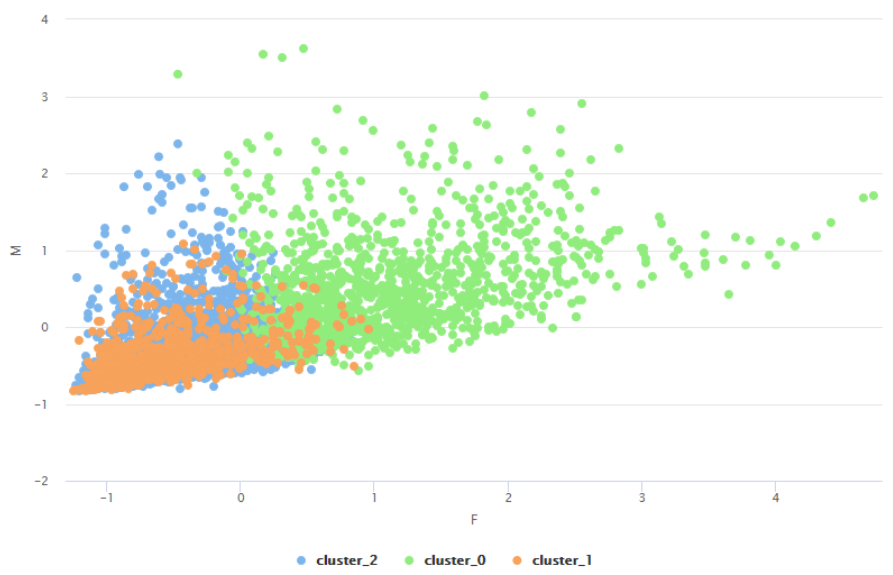
Η τρίτη συστάδα, παρουσιάζει χαμηλά κεντροειδές σε όλες τις διαστάσεις, και κάτω της μέσης τιμής.



Γράφημα 16: Κεντροειδές απεικόνιση διαστάσεων RFMV για τις συστάδες

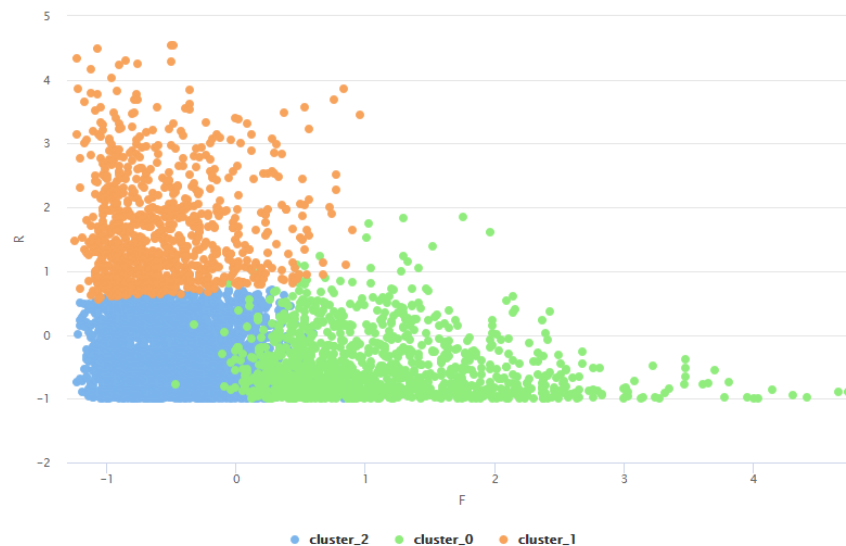
Στο παρακάτω διάγραμμα διασποράς (γράφημα 17) φαίνονται οι πυκνότητες των συστάδων ανάλογα με τις τιμές που έχουν στις διαστάσεις F και M.

Η πρώτη συστάδα (πράσινο χρώμα / cluster_0) εμφανίζει μεγάλη συγκέντρωση στην μέση τιμή και άνω της κανονικοποιημένης κλίμακας για την χρηματική αξία των αγορών και της συχνότητας αυτών, αλλά αρκετά παραδείγματα εκτείνονται μέχρι και τις ακραίες τιμές. Οι υπόλοιπες 2 συστάδες συγκεντρώνεται ως επί το πλείστον κάτω της μέσης τιμής.



Γράφημα 17: Διάγραμμα διασποράς συστάδων ως προς το F και το M

Στο 2^ο διάγραμμα διασποράς (γράφημα 18) διακρίνεται η συγκέντρωση που έχουν οι συστάδες ανάλογα τις τιμές των παραδειγμάτων για τις διαστάσεις R και F. Όπως και στο αρχικό μοντέλο, η τρίτη συστάδα (μπλέ χρώμα / cluster_2) αν και χωρίς πολλές επισκέψεις αγοράς, έχει τις από τις πιο πρόσφατες επισκέψεις, μαζί με την πρώτη συστάδα (πράσινο χρώμα / cluster_0) με τη διαφορά ότι αυτή η συστάδα έχει και την μεγαλύτερη συχνότητα αγορών. Η δεύτερη συστάδα (κόκκινο χρώμα / cluster_1) έχει να επισκεφθεί κάποιο κατάστημα αρκετό καιρό και η συχνότητα αγορών είναι επίσης χαμηλή.



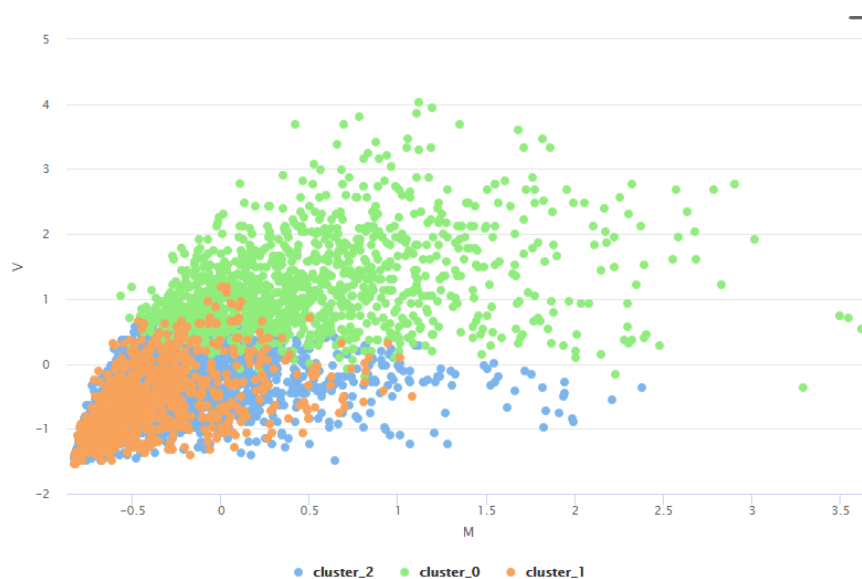
Γράφημα 18: Διάγραμμα διασποράς συστάδων ως προς το R και το F

Σε αυτό το διάγραμμα διασποράς (γράφημα 19) φαίνεται η καινούργια διάσταση του μοντέλου, η ποικιλία των προϊόντων που αγόρασαν οι πελάτες (V), ως προς την συχνότητα των αγορών τους. Όσο μεγαλύτερη η συχνότητα, τόσο μεγαλύτερη και η ποικιλία. Συγκεκριμένα, η πρώτη συστάδα (πράσινο χρώμα / cluster_0) που έχει τις μεγαλύτερες τιμές F, έχει και τις μεγαλύτερες τιμές V. Σημασία έχει όμως πως και οι άλλες 2 συστάδες συγκεντρώνονται κοντά στη μέση τιμή με αρκετά παραδείγματα να έχουν αγοράσει ποικιλία προϊόντων πάνω από τη μέση τιμή της V.



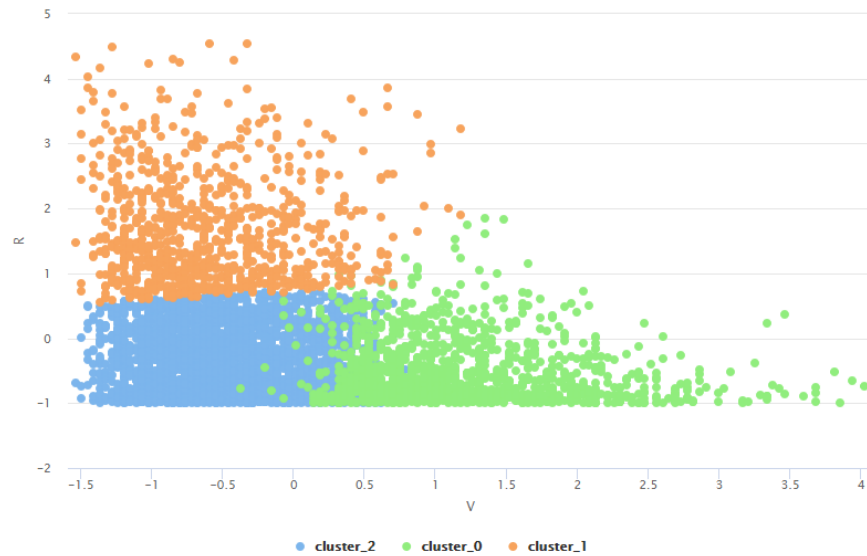
Γράφημα 19: Διάγραμμα διασποράς συστάδων ως προς το V και το F

Το 4^ο διάγραμμα διασποράς (γράφημα 20) του μοντέλου RFMV, απεικονίζει την ένταση που έχουν οι συστάδες ως προς την χρηματική αξία και την ποικιλία προϊόντων των αγορών τους. Η δεύτερη και τρίτη συστάδα (cluster_1 και cluster_2 αντίστοιχα) συγκεντρώνονται σε τιμές κάτω των μέσων, τόσο για τη διάσταση M όσο και για την V, με την τρίτη συστάδα να έχει κάποια παραδείγματα με αυξημένη χρηματική αξία αγορών. Η πρώτη συστάδα ωστόσο (cluster_0) έχει περισσότερα παραδείγματα υψηλών τιμών M, και εκτείνεται μέχρι τις ακραίες τιμές τις V.



Γράφημα 20: Διάγραμμα διασποράς συστάδων ως προς το V και το M

Στο τελευταίο διάγραμμα διασποράς (γράφημα 21) διακρίνεται η συγκέντρωση που έχουν οι συστάδες ως προς την επικαιρότητα των αγορών των πελατών και της ποικιλίας προϊόντων που αγόρασαν. Η δεύτερη συστάδα (πορτοκαλί χρώμα / cluster_1) με την τελευταία αγορά των πελατών να είναι παλαιότερη από τις άλλες 2 συστάδες, εμφανίζει χαμηλή ένταση στην ποικιλία προϊόντων παρεμφερή με την τρίτη συστάδα (μπλε χρώμα / cluster_2). Οι πελάτες της πρώτης συστάδας παρουσιάζουν τις πιο πρόσφατες αγορές και την μεγαλύτερη ποικιλία προϊόντων.



Γράφημα 21: Διάγραμμα διασποράς συστάδων ως προς το R και το V

Στην παρακάτω εικόνα (εικόνα 16) γίνεται παρουσίαση των συστάδων ως προς του όγκου τους και των διαστάσεών τους για το νέο μοντέλο RFMV.

Η πρώτη συστάδα (cluster_0), όγκου 1.230 πελατών έχει κατά μέσο όρο 99% μεγαλύτερες τιμές ως προς την συχνότητα επισκέψεων (F) από τις υπόλοιπες συστάδες, χρηματική αξία αγορών (M) κατά μέσο όρο 88% μεγαλύτερη, και ποικιλία προϊόντων που αγοράστηκαν (V) κατά μέσο όρο 84% περισσότερα από τις άλλες 2 συστάδες. Αν και δεν απεικονίζεται στην εικόνα λόγω περιορισμού του λογισμικού RapidMiner Studio, αυτή η συστάδα έχει και τις πιο πρόσφατες αγορές (R) σε σχέση με τις υπόλοιπες 2.

Η δεύτερη συστάδα (cluster_1), όγκου 847 πελατών, είναι η συστάδα με τις πιο παλαιές επισκέψεις, κατά μέσο όρο 172% μεγαλύτερο R, αλλά και την μικρότερη συχνότητα επισκέψεων με μέσο όρο 41%, και τη μικρότερη χρηματική αξία αγορών με μέσο όρο 38% χαμηλότερη από τις υπόλοιπες συστάδες. Το V είναι επίσης χαμηλό όπως και στην τρίτη συστάδα.

Η τρίτη συστάδα αποτελείται με πάνω από τους μισούς πελάτες του δείγματος και στο νέο μοντέλο, όγκου 2,695 πελατών. Σε αυτή τη συστάδα βρίσκονται πελάτες που έχουν πραγματοποιήσει σχετικά πρόσφατα αγορά, με μέσο όρο 32% μικρότερο R από τις υπόλοιπες συστάδες. Η χρηματική αξία και η συχνότητα των αγορών τους είναι κατά μέσο όρο περίπου 30% μικρότερες από τις άλλες 2 συστάδες.

Number of Clusters: 3

Cluster 0

1,230

F is on average 99.02% larger, M is on average 88.53% larger, V is on average 84.21% larger

Cluster 1

847

R is on average 172.81% larger, F is on average 41.00% smaller, M is on average 37.98% smaller

Cluster 2

2,695

F is on average 32.31% smaller, R is on average 31.61% smaller, M is on average 28.47% smaller

Εικόνα 16: Συστάδες μοντέλου RFMV

4.2.8 Ερμηνεία Αποτελεσμάτων RFMV

Η άμεση σύγκριση της RFM και της RFMV και η αξία που δίνει θα έπρεπε να υποστηριχθεί με επιπλέον δεδομένα από τα νοικοκυριά. Επειδή δεν είναι ρεαλιστικό να γίνει εις βάθος συνέντευξη με 5000 νοικοκυριά και επειδή δεν είναι δυνατή η πρόσβαση σε αυτά, προτιμήθηκε να διερευνηθούν τα αποτελέσματα της RFMV συγκριτικά με αυτά της RFM.

Όπως και στο αρχικό μοντέλο, και τώρα οι πελάτες συσταδοποιήθηκαν σε 3 συστάδες, σχεδόν ίδιου μεγέθους.

Με τη νέα διάσταση της ποικιλίας των προϊόντων που αγόρασαν οι πελάτες επιτεύχθηκε η προσθήκη μεγαλύτερης αξίας σε όλες τις συστάδες για την καλύτερη κατανόηση της αγοραστικής τους συμπεριφοράς, που είναι και ένας από τους στόχους αυτής της έρευνας. Η διάσταση V επιβεβαίωσε την σημαντικότητα της πρώτης και τρίτης συστάδας (δεύτερη και Τρίτη του αρχικού μοντέλου RFM) για την βιωσιμότητα και την σταθερή λειτουργία της επιχείρησης.

Η πρώτη και κύρια διαπίστωση είναι πως οι συχνότητα αγορών είναι άμεσα συσχετισμένη με την ποικιλία προϊόντων που αγοράζουν οι πελάτες. Επίσης, μεγαλύτερη δαπάνη ισοδυναμεί και με μεγαλύτερη ποικιλία προϊόντων ως ένα βαθμό.

Οι πελάτες της πρώτης συστάδας, όπως και οι πελάτες της δεύτερη συστάδας από το αρχικό μοντέλο RFM, εμφανίζουν τις υψηλότερες τιμές σε όλες τις διαστάσεις. Έχουν αγοράσει πρόσφατα, κάνουν συχνές αγορές, αγοράζουν μεγάλη ποικιλία προϊόντων και δαπανούν τα περισσότερα χρήματα στην επιχείρηση. Παίρνοντας υπόψη και τη χρονική διάρκεια των δεδομένων όπου είναι 2 χρόνια, αυτοί οι πελάτες δείχνουν ένα βαθμό πιστότητας προς την επιχείρηση και θα πρέπει να γίνουν ενέργειες για τη διατήρηση αυτής. Είναι η ιδανική συστάδα ως προς τις διαστάσεις RFMV.

Όπως και στο πρώτο μοντέλο, οι πελάτες που απαρτίζουν την δεύτερη συστάδα (πρώτη συστάδα μοντέλου RFM), φαίνεται να έχουν σταματήσει να είναι πελάτες της εταιρείας ή να έχουν ολοκληρώσει το έργο που είχαν αναλάβει, καθώς η τελευταία τους επίσκεψη είναι αρκετά παλαιά. Εμφανίζουν επίσης ένα χαρακτήρα οπορτουνιστικό καθώς δεν είχαν αρκετές επισκέψεις, ούτε δαπανούσαν αρκετά χρήματα. Η ποικιλία προϊόντων επιβεβαιώνει αυτή την συμπεριφορά καθώς και αυτή είναι η χαμηλότερη σε σχέση με τις υπόλοιπες 2 συστάδες. Θεωρείται πως οι αγορές που έκαναν είναι συγκεκριμένων προϊόντων όταν ήταν σε προσφορά, ή είχαν ανάγκη να αγοράσουν κάτι συγκεκριμένο εκείνη την στιγμή.

Η τρίτη και μεγαλύτερη συστάδα, με όμοια χαρακτηριστικά της τρίτης συστάδας και από το αρχικό μοντέλο RFM, αποτελείται από πιστούς πελάτες που αγοράζουν σταθερά από την επιχείρηση, έχουν σημαντική ποσότητα ποικιλίας στα προϊόντα που αγοράζουν, τα χρήματα που

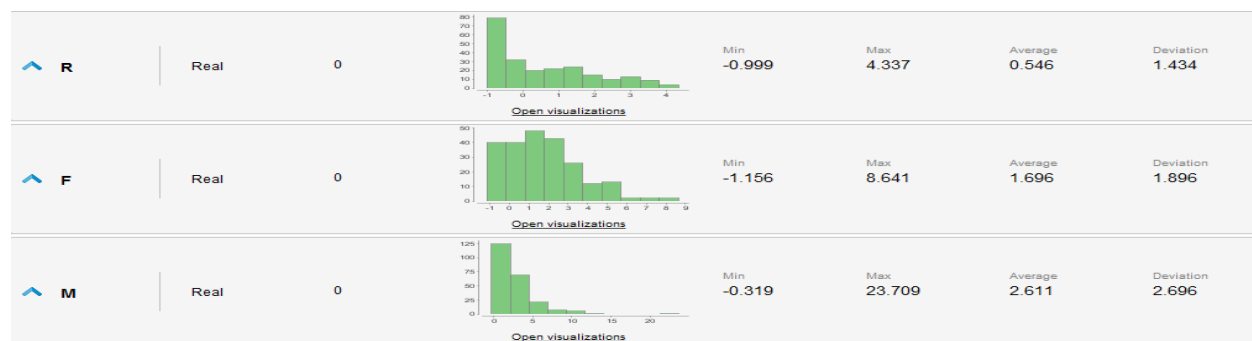
δαπανούν είναι αξιοσημείωτα αν παρθεί υπόψη ο όγκος αυτή της συστάδας και η τελευταία επίσκεψη ήταν πρόσφατη. Αυτή η συστάδα φέρει τη μεγαλύτερη και τη σταθερότερη χρηματοροή στην επιχείρηση και είναι η πιο σημαντική για την σταθερή της λειτουργία.

4.2.9 Μελέτη Ακραίων Τιμών

Και στα 2 μοντέλα, οι ακραίες τιμές είχαν απομονωθεί για να διερευνηθεί η αξία τους σαν ξεχωριστή συστάδα. Η συστάδα των ακραίων τιμών αποτελείται από 223 πελάτες στο μοντέλο RFM και από 228 στο μοντέλο RFMV, όπου είναι περίπου το 4% του συνολικού δείγματος και στις 2 περιπτώσεις.

Στις εικόνες 17 και 18 απεικονίζονται τα περιγραφικά στατιστικά των ακραίων τιμών ξεχωριστά για το κάθε μοντέλο. Οι τιμές για τις διαστάσεις R, F, M και V έχουν κανονικοποιηθεί και στα 2 μοντέλα. Οι μεγαλύτερες διαφορές φαίνονται στην συχνότητα αγορών και στην αξία αυτών. Όλοι οι μέσοι όροι είναι αρκετά αυξημένοι αλλά την μεγαλύτερη διαφορά έχει η χρηματική αξία στην ελάχιστη τιμή και ειδικά στην μέγιστη τιμή. Επίσης, η απόκλιση των παραδειγμάτων σε όλες τις διαστάσεις είναι αρκετά αυξημένες.

Εκ πρώτης όψεως, οι πελάτες αυτής της συστάδας δεν έχουν έρθει τόσο πρόσφατα για αγορά στην επιχείρηση με μέσο όρο 0,6 και στα 2 μοντέλα, η συχνότητα τους είναι καλή με μέσο όρο περίπου 1,550 και στα 2 μοντέλα. Ο μέσος όρος της χρηματικής αξίας των αγορών είναι περίπου 2,6, σημαντικά αυξημένος από τις υπόλοιπες συστάδες. Η ποικιλία των προϊόντων παίρνει και αυτή καλές τιμές καθώς ο μέσος όρος είναι στο 1,340.

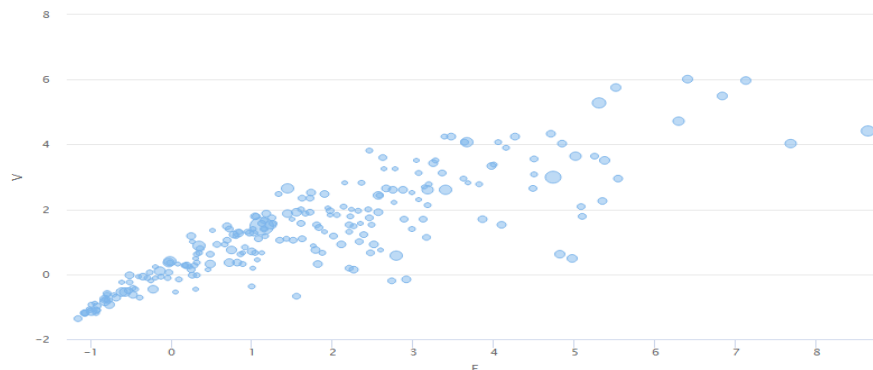


Εικόνα 17: Περιγραφικά στατιστικά ακραίων τιμών μοντέλου RFM



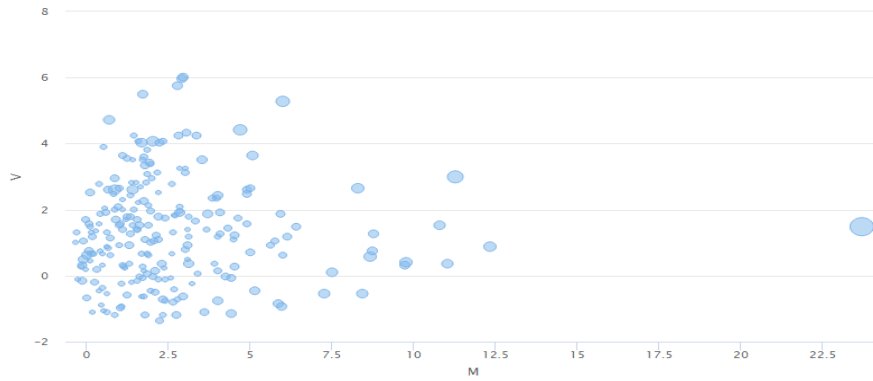
Εικόνα 18: Περιγραφικά στατιστικά ακραίων τιμών μοντέλου RFMV

Το μοντέλο RFMV επιβεβαιώνεται και για τις ακραίες τιμές ως προς το συσχετισμό της συχνότητας αγορών με την ποικιλία προϊόντων που αγοράστηκαν, όπως φαίνεται και από το διάγραμμα διασποράς (γράφημα 22), καθώς όσο μεγαλύτερη η συχνότητα, τόσο μεγαλύτερη και η ποικιλία.



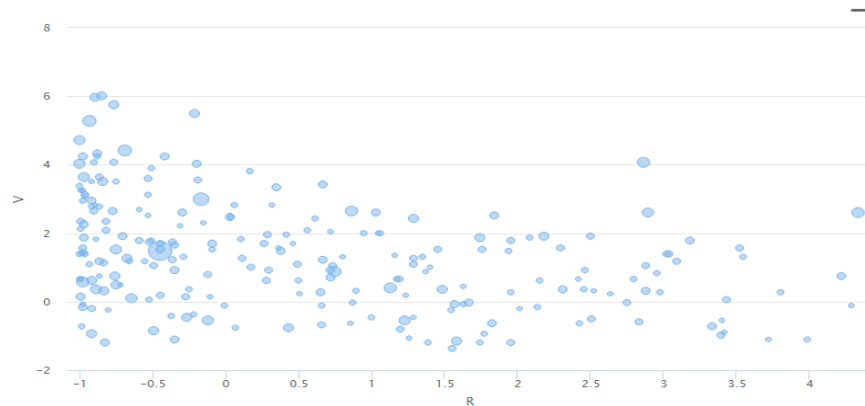
Γράφημα 22: Διάγραμμα διασποράς ακραίων τιμών ως προς το V και το F

Αναφορικά με την χρηματική αξία των αγορών και την ποικιλία προϊόντων που αγοράστηκαν, οι τιμές των παραδειγμάτων συγκεντρώνονται σε θετικές τιμές πάνω από τον μέσο όρο για τη διάσταση M (με μια αρκετά ακραία τιμή), αλλά εκτείνονται σε όλη τον άξονα της διάστασης V, όπως φαίνεται και από το γράφημα 23.



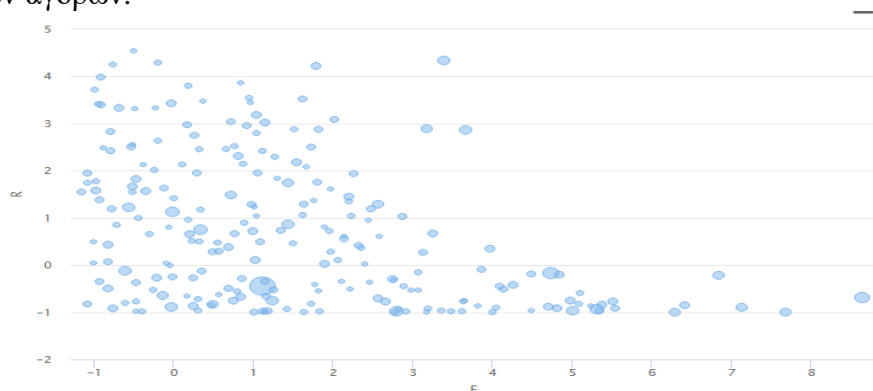
Γράφημα 23: Διάγραμμα διασποράς ακραίων τιμών ως προς το V και το M

Όσον αφορά την ποικιλία και την επικαιρότητα αγορών (γράφημα 24), οι τιμές των παραδειγμάτων φαίνονται διάσπαρτες και στις 2 διαστάσεις.



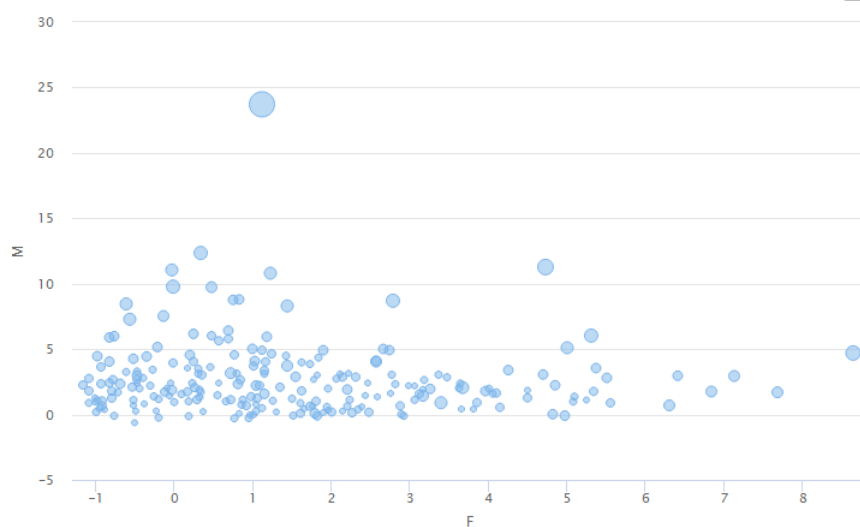
Γράφημα 24: Διάγραμμα διασποράς ακραίων τιμών ως προς το V και το R

Από το γράφημα 25 φαίνεται πως όσο παλαιότερη η επίσκεψη, τόσο μικρότερη είναι και η συχνότητα των αγορών.



Γράφημα 25: Διάγραμμα διασποράς ακραίων τιμών ως προς το R και το F

Η χρηματική αξία είναι αρκετά αυξημένη από τις υπόλοιπες συστάδες σε όλα τα παραδείγματα αλλά δεν επηρεάζεται τόσο πολύ από την συχνότητα αγορών.



Γράφημα 26: Διάγραμμα διασποράς ακραίων τιμών ως προς το M και το F

5 Συζήτηση - Συμπεράσματα

Σε αυτό το κεφάλαιο θα γίνει συνοψισμός των αποτελεσμάτων της εμπειρικής μελέτης, θα γίνει παρουσίαση των αποτελεσμάτων των στόχων που είχαν τεθεί στο πρώτο κεφάλαιο, θα παρουσιαστούν τα κύρια συμπεράσματα από την δημιουργία της τμηματοποίησης των πελατών του δείγματος, θα προσφερθούν προτάσεις για τη κάθε συστάδα και θα γίνει προσπάθεια σύγκρισης με άλλες μελέτες. Τέλος, θα αναφερθούν περιορισμοί που υπήρξαν στην έρευνα και προτάσεις για μελλοντική μελέτη.

5.1 Αποτελέσματα

Στην αρχή της έρευνας είχαν τεθεί 3 στόχοι:

1. Να μελετηθεί η αξία των πελατών της εταιρείας με χρήση της ανάλυσης RFM
2. Να γίνει κριτική της τεχνικής τμηματοποίησης RFM και των αποτελεσμάτων που δημιουργεί
3. Να ενισχυθεί η τεχνική τμηματοποίησης RFM με την προσθήκη καινούργιας διάστασης

Από αυτή την ανάλυση RFM, οι πελάτες του δείγματος τμηματοποιήθηκαν με επιτυχία ως προς τις διαστάσεις της επικαιρότητας των επισκέψεων τους, της συχνότητας αυτών και της χρηματικής αξίας που δαπανούσαν στις αγορές τους. Η αξία του κάθε πελάτη είναι εμφανής, και έχει συσταδοποιηθεί μαζί με άλλους πελάτες που μοιράζονται παρόμοια χαρακτηριστικά αγορών. Οι υπεύθυνοι μάρκετινγκ είναι τώρα σε θέση να σχεδιάσουν τις κατάλληλες ενέργειες μάρκετινγκ για την κάθε συστάδα.

Χρησιμοποιώντας την ανάλυση RFM σαν εργαλείο τμηματοποίησης σε μεγάλες επιχειρήσεις, μπορεί να προσφέρει ωφέλιμες πληροφορίες και βαθύτερη γνώση των πελατών της για την καλύτερη διαχείριση τους και στόχευση μέσω τακτικών μάρκετινγκ. Ο δεύτερος στόχος επιτεύχθηκε καθώς ένα μεγάλο δείγμα πελατών, μπόρεσε να τμηματοποιηθεί και να αναδειχθεί η αξία του βάσει των αγοραστικών του προτιμήσεων και αγοραστικών συμπεριφορών, μέσω της επιτυχής εφαρμογής της τεχνικής RFM. Η εταιρεία είναι σε θέση να εφαρμόσει το ίδιο μοντέλο σε μεγαλύτερο αριθμό πελατών της, οποιαδήποτε χρονική στιγμή, βάζοντας την σε πλεονεκτική θέση έχοντας στην διάθεσή της ένα τμηματοποιημένο πελατολόγιο και να πάρει καλύτερες στρατηγικές αποφάσεις για την διαχείριση του κάθε τμήματος πελατών.

Με την προσθήκη της τέταρτης διάστασης, συγκεκριμένα της ποικιλίας των προϊόντων που αγόρασαν οι πελάτες, επιτεύχθηκε και ο τρίτος στόχος της έρευνας προσθέτοντας επιπλέον αξία σε όλες τις συστάδες που δημιουργήθηκαν καθώς τώρα η εταιρεία γνωρίζει καλύτερα την συμπεριφορά και τις προτιμήσεις των πελατών. Επιπλέον, η καλύτερη συστάδα βάσει των διαστάσεων RFM επιβεβαιώθηκε καθώς ήταν αυτοί όπου είχαν αγοράσει τη μεγαλύτερη ποικιλία προϊόντων. Η προσθήκη διαστάσεων στη παραδοσιακή ανάλυση RFM προσδίδει μεγαλύτερη αξία στην τμηματοποίηση πελατών. Ο Kotler (2010) αναφέρει πως οι καταναλωτές εξελίσσονται από consumer σε prosumer, παράγοντας και καταναλώνοντας ταυτόχρονα για διάφορους λόγους όπως οικονομία και αίσθηση ευχαρίστησης από την παραγωγή. Η εταιρεία της έρευνας χαρακτηρίζεται για την προσφορά προϊόντων ιδιοκατασκευής, και με την συσταδοποίηση των πελατών, εμφανίζεται ένα τμήμα όπου αγοράζει μεγάλη ποικιλία προϊόντων αρκετά συχνά και δαπανά σημαντική χρηματική αξία. Αυτοί οι πελάτες μπορούν να χαρακτηριστούν ως prosumer, και μάλιστα ειδικεύεται η αναφορά του Kotler (2010), καθώς βάσει των αποτελεσμάτων της παρούσας έρευνας, είναι εμφανής και η πληθώρα των προϊόντων που αγοράζουν.

Με την χρήση των μοντέλων τμηματοποίησης RFM και της επέκτασής του RFMV, επιτεύχθηκε η δημιουργία 3 συστάδων από το σύνολο των 5.000 πελατών.

Η πρώτη συστάδα του μοντέλου RFM, που παρουσιάζει παρόμοια χαρακτηριστικά με την δεύτερη συστάδα του μοντέλου RFMV, αποτελείται από περίπου 860 πελάτες οι οποίοι δεν έχουν πραγματοποιήσει αγορά για μεγάλο χρονικό διάστημα. Η ποικιλία των προϊόντων που αγόρασαν, η χρηματική αξία των αγορών τους και η συχνότητα αυτών είναι επίσης οι χαμηλότερες του δείγματος. Η συστάδα αυτή έχει τη μικρότερη αξία στην επιχείρηση. Αυτοί οι πελάτες θα μπορούσαν να χαρακτηριστούν ως ομορτοουνιστικοί πελάτες, οι οποίοι έκαναν κάποιες αγορές σε περιόδους προσφορών, ή είχαν αναλάβει κάποιο έργο ιδιοκατασκευής, όπου με την ολοκλήρωση του, σταμάτησαν να αγοράζουν από την εταιρεία. Παίρνοντας υπόψη τις χαμηλές τιμές σε όλες τις διαστάσεις, οι συγκεκριμένοι πελάτες θεωρείται πως είχαν αναλάβει κάποιο έργο βελτίωσης σπιτιού ή ιδιοκατασκευής μικρής κλίμακας. Σίγουρα θα πρέπει να παρθεί υπόψη και ο ανταγωνισμός, που μπορεί να έχει κερδίσει αυτούς τους πελάτες.

Αν η επιχείρηση αποφάσιζε να προσπαθήσει να επαναφέρει αυτούς τους πελάτες θα πρέπει να γίνουν ενέργειες μάρκετινγκ “επανενεργοποίησης” πελατών. Προτείνονται τα παρακάτω:

1. Προώθηση εκπτώτικου κωδικού σημαντικής αξίας για την επόμενη τους αγορά
2. Προσφορά μίας δωρεάν διαβούλευσης με επαγγελματία αρχιτέκτονα ή σχεδιαστή εσωτερικού χώρου
3. Επικοινωνία άμεσου μάρκετινγκ με θέματα για βελτίωση σπιτιού και ιδέες ιδιοκατασκευής
4. Επικοινωνία άμεσου μάρκετινγκ για το πρόγραμμα επιβράβευσης που παρέχει η εταιρεία και τα οφέλη συμμετοχής σε αυτό

Η δεύτερη συστάδα, η οποία είναι παρόμοια με την πρώτη συστάδα από την επέκταση RFMV, παρουσιάζει τις ιδανικές τιμές σε όλες τις διαστάσεις. Αυτοί οι 1.160 πελάτες περίπου, έχουν αγοράσει πολύ πρόσφατα, έχουν δαπανήσει τα περισσότερα χρήματα, αγοράζουν πάρα πολύ συχνά από την εταιρεία, αλλά έχουν αγοράσει και την μεγαλύτερη ποικιλία προϊόντων. Σύμφωνα με την ανάλυση RFM, αυτή η συστάδα έχει την μεγαλύτερη αξία για την επιχείρηση. Παίρνοντας υπόψη τη διετή περίοδο παρατήρησης της έρευνας και της υψηλής συχνότητας αυτών των πελατών, μπορεί να θεωρηθεί πως είναι αρκετά πιστοί στην εταιρεία. Λόγω της υψηλής χρηματικής αξίας των αγορών τους και της μεγάλης ποικιλίας των προϊόντων που αγοράζουν, τα έργα που αναλαμβάνουν είναι μεγάλης κλίμακας (ανακαίνιση σπιτιού) ή/και αναλαμβάνουν πολλά και διαφορετικά έργα ιδιοκατασκευής, όπως η δημιουργία επίπλων, ή ανακαίνιση δωματίων. Για τη παρούσα συστάδα, προτείνονται οι παρακάτω ενέργειες μάρκετινγκ για την διατήρηση τους ως πελάτες:

1. Αποκλειστικές προσφορές και εκπτωτικοί κωδικοί εκτός από αυτά που ισχύουν για το ευρύ πελατειακό κοινό της εταιρείας
2. Συμμετοχή σε πρόγραμμα εξοικονόμησης με σταθερό ποσοστό έκπτωσης σε κάθε τους αγοράς. Στο πρόγραμμα εξοικονόμησης χρειάζεται μια και μοναδική πληρωμή κάθε χρόνο
3. Επικοινωνία νέων προϊόντων (εργαλείων/υλικών) μέσω άμεσου μάρκετινγκ πριν την διαθεσιμότητά τους και δυνατότητα προ-παραγγελίας με μικρή έκπτωση
4. Επικοινωνία νέων προσφορών πριν ξεκινήσουν μέσω άμεσου μάρκετινγκ. Παράδειγμα μπορεί να είναι η πληροφόρηση των προϊόντων που θα είναι σε έκπτωση την ημέρα της Black Friday

Τέλος, η τρίτη και μεγαλύτερη συστάδα, η οποία είναι παρόμοια επίσης της τρίτης από το μοντέλο RFMV, ίσως είναι η σημαντικότερη συστάδα για την εταιρεία για τους εξής λόγους.

- Αποτελείται από πάνω από τους μισούς πελάτες του δείγματος (περίπου 2.740)
- Αποτελείται από πελάτες οι οποίοι έχουν κάνει πρόσφατα αγορά
- Οι ποικιλία των προϊόντων που αγοράζουν, η χρηματική αξία που δαπανούν και η συχνότητα των αγορών τους κυμαίνεται σε καλά επίπεδα, όχι μακριά από τη μέση τιμή των διαστάσεων
- Δείχνουν σταθερή παρουσία και τα 2 χρόνια που είναι η περίοδος παρατήρησης της έρευνας

Αυτή η συστάδα φέρει την μεγαλύτερη και σταθερότερη χρηματοροή στην εταιρεία, και είναι αναγκαία στην βιωσιμότητά της. Θα πρέπει να εφαρμοσθούν τακτικές μάρκετινγκ “διατήρησης” πελατών, αλλά επίσης, η εταιρεία θα πρέπει να επενδύσει σημαντικά και στην καλή εξυπηρέτηση πελατών για την καλύτερη δυνατή αγοραστική εμπειρία, πριν και μετά την αγορά, καθώς σε κλίμακα, αυτό είναι και το ευρύτερο κοινό της εταιρείας. Όσων αφορά τις ενέργειες μάρκετινγκ, προτείνονται τα παρακάτω:

1. Συχνή επικοινωνία μέσω άμεσου μάρκετινγκ του προγράμματος επιβράβευσης και του προγράμματος πιστότητας που παρέχει η εταιρεία, και τα οφέλη συμμετοχής σε αυτό
2. Δημιουργία και επικοινωνία προγράμματος σύστασης φίλων όπου μετά την επιτυχή αγορά αυτών, ο κύριος πελάτης θα έχει κάποια συγκεκριμένη επιβράβευση (π.χ. μικρή έκπτωση στην επόμενη του αγορά)
3. Συχνές και διαφορετικές καμπάνιες με προσφορές σε συγκεκριμένους τομείς προϊόντων (π.χ. προσφορές σε είδη κουζίνας)
4. Διαγωνισμοί στα κοινωνικά δίκτυα της εταιρείας όπου πελάτες θα παρουσιάζουν τις δημιουργίες ιδιοκατασκευής τους και βάσει των “like” που θα έχουν από άλλους ακόλουθους των σελίδων social media της εταιρείας, θα κερδίζουν συγκεκριμένα βραβεία (π.χ. δωρεάν υλικά/εξοπλισμός για το επόμενο τους έργο ιδιοκατασκευής μέχρι ένα συγκεκριμένο ποσό)

5.2 Σύγκριση Αποτελεσμάτων με Άλλες Έρευνες

Ο Hughes (2000) ήταν πρωτοπόρος στη δημιουργία ενεργειών μάρκετινγκ παίρνοντας υπόψη τη βάση δεδομένων και έχει αναδείξει την αξία της ανάλυσης RFM μέσα από τις πολλές του έρευνες. Η δικιά του προσέγγιση ήταν η τμηματοποίηση του δείγματος σε 5 ίσα μέρη ανάλογα την επικαιρότητα των αγορών (R), την συχνότητα αγορών (F) και της χρηματικής αξίας που δαπανούσαν οι πελάτες (M), δημιουργώντας έτσι το κατώτερο 20% που αποτελείται από πελάτες που δεν ψωνίζουν συχνά, δεν έχουν πραγματοποιήσει αγορά πρόσφατα αλλά και η χρηματική αξία είναι χαμηλή. Την ίδια στιγμή, δημιουργείται και το ανώτερο 20% πελατών που αγοράζει συχνά, έχει αγοράσει πρόσφατα και δαπανά μεγάλη αξία χρημάτων στην επιχείρηση. Σαφώς και δημιουργούνται και τα υπόλοιπα 3 τμήματα των 20% στο ενδιάμεσο. Αν και στη παρούσα εργασία δεν ακολουθήθηκε αυτή η προσέγγιση, με επιτυχία έγινε η τμηματοποίηση των πελατών του δείγματος και αναδείχθηκαν 3 πολύ ξεχωριστές συστάδες, όπου υπάρχει η “ιδανική” συστάδα βάση της RFM, αλλά και αυτή που έχει τη λιγότερη αξία.

Οι Zhang et al. (2014) ανέφεραν την ανεπάρκεια της παραδοσιακής τμηματοποίησης μέσω της ανάλυσης RFM για την ακριβή πρόγνωση μελλοντικής συμπεριφοράς των πελατών και ανταπόκρισής τους σε ενέργειες μάρκετινγκ, και δημιούργησαν μια νέα επέκταση, προσθέτοντας τη διάσταση C. Η C, ή αλλιώς συσσωμάτωση – clumpiness, παίρνει υπόψη της την συσσωμάτωση που δημιουργείται ανάμεσα στους πελάτες σε αρκετές μεταβλητές όπως στα δημογραφικά, στις αγορές που πραγματοποιούν, στις ενέργειες μάρκετινγκ που ανταποκρίνονται, κλπ. Με αυτή τους την επέκταση, υποστηρίζουν την προσθήκη επιπλέον διαστάσεων στην ανάλυση RFM καθώς από τα αποτελέσματά τους, αναδείχθηκε καλύτερη η διαχρονική αξία του πελάτη (CLV – Customer Lifetime Value). Μέσω της επέκτασης που δημιουργήθηκε στη παρούσα έρευνα, έγινε καλύτερη σκιαγράφηση του προφίλ των πελατών της εταιρείας βελτίωσης σπιτιού και ιδιοκατασκευής, και προστέθηκε μεγαλύτερη αξία στις συστάδες.

Με τη σειρά τους οι Heldt et al. (2021) δημιούργησαν και αυτή μια νέα επέκταση της RFM στην έρευνά τους, την P ή αλλιώς προϊόν – Product, καθώς υποστηρίζουν και αυτή πως η παραδοσιακή ανάλυση RFM έχει κάποιους περιορισμούς. Προσθέτοντας της διάσταση της P, η ανάλυση RFMP δημιούργησε συστάδες πελατών βάση των προϊόντων που είχαν αγοράσει, αναδεικνύοντας έτσι την αξία που έχει το κάθε προϊόν στην εταιρεία. Στη παρούσα έρευνα, η διάσταση που προστέθηκε ήταν αυτή της ποικιλίας προϊόντων (V – Variety). Το αποτέλεσμα ήταν κυρίως η επιβεβαίωση της μεγάλης αξίας που έχει η ιδανική συστάδα RFM, καθώς εκτός από πρόσφατες αγορές και συχνές αγορές με μεγάλη αξία, οι πελάτες αγόραζαν και μεγάλη ποικιλία προϊόντων. Σε δεύτερη φάση, αυτή η καινούργια μεταβλητή μπορεί να χρησιμοποιηθεί για προώθηση μεγαλύτερη ποικιλίας προϊόντων σε αυτό το τμήμα πελατών καθώς είναι πιο πιθανό να προχωρήσουν σε αγορά.

5.3 Περιορισμοί Έρευνας

Ο πρώτος και κύριος περιορισμός της παρούσας έρευνας ήταν η παλαιότητα του δείγματος. Δεδομένου του δυναμικού και ταχέως μεταβαλλόμενου περιβάλλοντος στο οποίο ανταγωνίζονται οι περισσότεροι οργανισμοί (Kennerley & Neely, 2003), τα δεδομένα και η πελατειακή βάση της επιχείρησης σίγουρα έχουν αλλάξει. Για αυτό το λόγο, θα πρέπει να γίνεται τακτικά καινούργια ανάλυση και τμηματοποίηση, ώστε οι ενέργειες μάρκετινγκ, ή ακόμα και στρατηγικές ενέργειες, να έχουν καλύτερη απόδοση. Παρόλο που οι αγοραστικές συμπεριφορές και το πελατειακό καλάθι να έχουν αλλάξει ραγδαία με την πάροδο του χρόνου, το μοντέλο που κατασκευάστηκε για την τμηματοποίηση RFM και RFMV στη παρούσα έρευνα, μπορεί να χρησιμοποιηθεί και με πιο πρόσφατα δεδομένα. Με πολύ λίγες αλλαγές στις ρυθμίσεις, μπορεί να παρέχει στον ερευνητή την δυνατότητα τμηματοποίησης και ανάλυσης της πελατειακής βάσης οποιαδήποτε στιγμή το χρειαστεί.

Δεύτερος περιορισμός είναι το δείγμα πελατών που υπήρχε διαθέσιμο. Η επιχείρηση όπου εφαρμόστηκε το μοντέλο ανήκει στις 500 μεγαλύτερες επιχειρήσεις του κόσμου, και χαιρείται πολύ μεγάλης πελατειακής βάσης. Αν και τα αποτελέσματα του μοντέλου είναι έγκυρα για τα άτομα του δείγματος και τα αγοραστικό προφίλ των πελατών σκιαγραφήθηκαν με επιτυχία, σε μια έρευνα όπου το δείγμα πελατών είναι μεγαλύτερο, τότε τα αποτελέσματα θα είναι αρκετά διαφορετικά. Το μοντέλο ωστόσο, μπορεί να τροφοδοτηθεί με πολύ μεγαλύτερο όγκο δεδομένων πελατών και με πολύ μικρές και ελάχιστες αλλαγές στη διαδικασία ως προς τις ρυθμίσεις της συσταδοποίησης, είναι ικανό να αποδώσει ακριβέστερα αποτελέσματα για τους πελάτες της επιχείρησης.

5.4 Προτάσεις για Περαιτέρω Έρευνα

Πρώτιστος, προτείνεται καινούργια ανάλυση με πρόσφατα δεδομένα καθώς αυτά της παρούσης έρευνας είναι από τη περίοδο 2012 με 2014. Σημαντική και ενδιαφέρουσα θα είναι η παρατήρηση

της εξέλιξης των συστάδων και της σημερινής αξίας των πελατών της επιχείρησης συγκριτικά με το παρελθόν. Κυρίως όμως, σε οποιαδήποτε ενέργεια μάρκετινγκ θα θέλει να προχωρήσει η εταιρεία, η οποία ενέργεια χρειάζεται πληροφορίες σχετικά με την αξία των πελατών, τότε θα χρειαστεί και καινούργια ανάλυση RFM ή RFMV των πελατών ώστε τα αποτελέσματα να είναι επίκαιρα.

Η εταιρεία έχει καταστήματα τόσο σε επαρχιακές περιοχές όσο και σε μεγαλουπόλεις. Οι πελάτες του συγκεκριμένου δείγματος προέρχονται από διάφορες περιοχές και οι αγοραστικές συμπεριφορές τους είναι και αυτές αρκετά διαφορετικές, όπως παραδείγματος χάρη ένας πελάτης ο οποίος διαμένει σε μεγάλη οικία στην Λουϊζιάνα και θέλει να αναλάβει ο ίδιος την ανακαίνιση της κατοικίας του, υπάρχει μεγάλη πιθανότητα τα τετραγωνικά μέτρα να είναι πολύ μεγαλύτερα, ή/και τα υλικά που θα χρειαστεί να είναι διαφορετικά (π.χ. αντοχή σε ανεμοστρόβιλους), σε αντίθεση με ένα πελάτη που θέλει να κάνει ανακαίνιση στο διαμέρισμά του στη Νέα Υόρκη. Μια ανάλυση RFM η οποία θα έπαιρνε υπόψη της την τοποθεσία του καταστήματος θα άλλαζε σημαντικά τις συστάδες και θα πρόσδιε ακόμη περισσότερη αξία στην ανάλυση καθώς η κάθε τοπική συστάδα θα μπορούσε να στοχευθεί καλύτερα με ενέργειες μάρκετινγκ.

Τέλος, προτείνεται η προσθήκη διαφορετικών επιπλέον διαστάσεων στην ανάλυση RFM, όπως έγινε σε αυτή την έρευνα με την προσθήκη της V (Variety / Ποικιλία αγορασθέντων προϊόντων). Όπως προαναφέρθηκε στη δεύτερη πρόταση, η προσθήκη της τοποθεσίας θα μπορούσε να εμφανίσει την αξία των πελατών ανά πόλη ή πολιτεία. Άλλες προτάσεις είναι η προσθήκη της διάστασης του τομέα προϊόντος (π.χ. κουζίνα) ώστε να βρεθεί ο τομέας προϊόντων με την μεγαλύτερη αξία για την επιχείρηση ανάλογα με την συστάδα, ή η προσθήκη της διάστασης των τετραγωνικών μέτρων της κατοικίας ώστε να διαπιστωθεί η αξία των πελατών ανάλογα με το μέγεθος της κατοικίας τους. Οι πιθανοί συνδυασμοί είναι πολυάριθμοι και όλοι φέρουν επιπρόσθετη αξία σε μια ήδη αποτελεσματική και με σημαντική αξία ανάλυση, την RFM.

6 Βιβλιογραφία

1. Aeron, H., Kumar, A., & Moorthy, J. (2012). Data mining framework for customer lifetime value-based segmentation. *Journal of Database Marketing & Customer Strategy Management*, 19, 17–30.
2. Baker, M. J., & Hart, S. (2008). *The Marketing Book, Sixth Edition*. Oxford: Butterworth - Heinemann.
3. Beane, T., & Ennis, D. (1987). Market Segmentation: A Review. *European Journal of Marketing*, 21, 20-42.
4. Blattberg, R. C., Kim, B.-D., & Neslin, S. A. (2008). *Database Marketing: Analyzing and Managing Customers*. New York: Springer.
5. Brocke, J. v., Maedche, A., & Hevner, A. (2020). Introduction to Design Science Research. *Design Science Research. Cases*, 1-13.
6. Changa, H.-C., & Tsai, H.-P. (2011). Group RFM analysis as a novel framework to discover better customer consumption behavior. *Expert Systems with Applications*, 38(12), 14499-14513.
7. Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C., & Wirth, R. (2000). *CRISP-DM 1.0 - Step-by-step data mining guide*. SPSS, CRISP-DM Consortium.
8. Cooil, B., Aksoy, L., & Keiningham, T. (2008). Approaches to Customer Segmentation. *Journal of Relationship*, 6, 9-39.
9. Coussement, K., Bossche, F. V., & Bock, K. D. (2014). Data accuracy's impact on segmentation performance: Benchmarking RFM analysis, logistic regression, and decision trees. *Journal of Business Research*, 67, 2751-2758.
10. Cross, G., & Thompson, W. (2008). Understanding Your Customer: Segmentation Techniques for Gaining Customer Insight and Predicting Risk in the Telecom Industry . *Data Mining and Predictive Modeling*, 1-14.
11. Davenport, T. H. (2007). *Competing on Analytics: The New Science of Winning*. Boston: Harvard Business Review Press.
12. Davenport, T. H. (2014). *Big Data at Work: Dispelling the myths uncovering the opportunities*. Boston: Harvard Business Review Press.
13. Dwyer, R. F. (1997). Customer lifetime valuation to support marketing decision making. *Journal of Direct Marketing*, 11, 6-13.
14. Fader, P. S., Hardie, B. G., & Lee, K. L. (2005). RFM and CLV: Using Iso-Value Curves for Customer Base Analysis. *Journal of Marketing Research*, 42, 415-430.

15. Fader, P., Hardie, B., & Lee, K. L. (2005). RFM and CLV: Using Iso-Value Curves for Customer Base Analysis. *Journal of Marketing Research*, 42, 415–430.
16. Fan, S., Lau, R. Y., & Zhao, J. L. (2015). Demystifying Big Data Analytics for Business Intelligence Through the Lens of Marketing Mix. *Big Data Research*, 2, 28–32.
17. Fotaki, G., Spruit, M., & Brinkkemper, S. (2014). Exploring Big Data Opportunities for Online Customer Segmentation. *International Journal of Business Intelligence Research*, 5, 58-75.
18. Gupta, S., & Zeithaml, V. (2006). Customer Metrics and Their Impact on Financial Performance. *Marketing Science*, 25, 718-739.
19. Hatton-Jones, S., & Teah, M. (2015). Case analysis of the do-it-yourself industry. *Asia Pacific Journal of Marketing and Logistics*, 27, 826-838.
20. Heldt, R., Silveira, G. S., & Luce, F. B. (2021). Predicting customer value per product: From RFM to RFM/P. *Journal of Business Research*, 127, 444-453.
21. Hevner, A., & Chatterjee, S. (2010). Design Science Research in Information Systems. *Design Science Research in Information Systems*, 22, 9-22.
22. Hevner, A., March, S., Park, J., & Ram, S. (2004). Design Science in Information Research. *MIS Quarterly*, 28, 75-105.
23. Hogan, J., Lemon, K., & Rust, R. (2002). Customer Equity Management: Charting New Directions for the Future of Marketing. *Journal of Service Research*, 5, 4-12.
24. Hughes, A. M. (2000). *Strategic Database Marketing*. New York: McGraw-Hill.
25. Hughes, A. M. (2005). Use RFM to boost your response rate. *Georgia World Congress Center* (pp. 1-72). Atlanta: Knowledge Base Marketing Inc.
26. Hughes, A. M., & Sellers, J. (2004). *RFM Migration Analysis: A new approach to a proven technique*. Ανάκτηση από Database Marketing Institute: <http://www.dbmarketing.com/articles/Art123.htm>
27. Hwang, H., Jung, T., & Suh, E. (2004). An LTV model and customer segmentation based on customer value: a case study on the wireless telecommunication industry. *Expert Systems with Applications*, 26, 181–188.
28. Johannesson, P., & Perjons, E. (2014). *An Introduction to Design Science*. Stockholm: Springer.
29. Jonker, J.-J., Piersma, N., & Poel, D. V. (2004). Joint optimization of customer segmentation and marketing policy to maximize long-term profitability. *Expert Systems with Applications*, 27, 159–168.

30. Kennerley, M., & Neely, A. (2003). Measuring Performance in a Changing Business Environment. *International Journal of Operations & Production Management*, 23, 213-229.
31. Khajvand, M., Zolfaghar, K., Ashoori, S., & Alizadeh, S. (2011). Estimating customer lifetime value based on RFM analysis of customer purchase behavior: case study. *Procedia Computer Science*, 3, 57-63.
32. Kotler, P. (2010). The Prosumer Movement. In: Blättel-Mink, B., Hellmann, KU. (eds) *Prosumer Revisited. VS Verlag für Sozialwissenschaften.*, 51-60.
33. Kotler, P., & Keller, K. L. (2016). *Marketing Management 15th Edition*. Essex: Pearson.
34. Kotler, P., Wong, V., Saunders, J., & Armstrong, G. (2005). *Principles of Marketing*. Essex: Pearson.
35. Kottler, P. (2003). *Marketing Management, Millenium Edition*. New Jersey: Pearson Custom Publishing.
36. Kumar, V., Venkatesan, R., Bohling, T., & Beckmann, D. (2008). The Power of CLV: Managing Customer Lifetime Value at IBM. *Marketing Science*, 27, 585–599.
37. Lamb, C. W., Hair, J. F., & McDaniel, C. (2011). *Marketing, 11th edition*. Mason: South-Western Cengage Learning.
38. Lemon, K., & Mark, T. (2008). Customer Lifetime Value as the Basis of Customer Segmentation. *Journal of Relationship Marketing*, 5, 55-69.
39. Lewis, M., Singh, V., & Fay, S. (2006). An Empirical Study of the Impact of Nonlinear Shipping and Handling Fees on Purchase Incidence and Expenditure Decisions. *Marketing Science*, 51-64.
40. March, S. T., & Smith, G. F. (1995). The Sciences of the Artificial Design and Natural Science Research on Information Technology. *Decision Support Systems*, 15, 251-266.
41. March, S., Hevner, A., & Ram, S. (2000). Research Commentary: An Agenda for Information Technology Research in Heterogeneous and Distributed Environments. *Informations System Research*, 11, 327-341.
42. Marcus, C. (1998). A practical yet meaningful approach to customer segmentation. *JOURNAL OF CONSUMER MARKETING*, 15, 494-504.
43. Massnick, F. (1997). *The Customer is CEO: How to Measure what Your Customers Wants – and Make Sure They Get It*. Chicago: Amacom.
44. Meredith, G., Schewe, C., & Karlovich, J. (2002). *Defining Marrkets, Definning Moments*. New York: Hungry Minds.

45. Monalisa, S., Nadya, P., & Novita, R. (2019). Analysis for Customer Lifetime Value Categorization with RFM Model. *Procedia Computer Science*, 161, 834-840.
46. Petersen, R. (2019, January 27). *6 Audience segmentation tool to find customers that need you most*. Retrieved from BarnRaisers: <https://barnraisersllc.com/2019/01/27/audience-segmentation-tools-find-customers/>
47. Rokeach, M. (1973). *The Nature of Human Values*. New York: The Free Press.
48. Rossi, P., Allenby, G., & McCulloch, R. (1996). The Value of Purchase History Data in Target Marketing. *Marketing Science*, 15, 321-340.
49. Rust, R. T., & Verhoef, P. C. (2005). Optimizing the Marketing Interventions Mix in Intermediate-Term CRM. *Marketing Science*, 24, 477-489.
50. Smith, W. (1956). PRODUCT DIFFERENTIATION AND MARKET SEGMENTATION AS ALTERNATIVE MARKETING STRATEGIES. *Journal of Marketing*, 21, 3-8.
51. Sun, S. (2009). An Analysis on the Conditions and Methods of Market Segmentation. *International Journal of Business & Management*, 4, 63-70.
52. Tsiptsis, K., & Chorianopoulos, A. (2009). *Data Mining Techniques in CRM: Inside Customer Segmentation*. West Sussex: Wiley.
53. Tynan, C., & Drayton, J. (1987). Market Segmentation. *Journal of Marketing Management*, 2, 301-335.
54. Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a Standard Process Model for Data Mining. *Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining*, (pp. 29-39).
55. Wolf, M., & McQuitty, S. (2011). Understanding the do-it-yourself consumer: DIY motivations and outcomes. *Academy of Marketing Science*, 1, 154-170.
56. Zhang, Y., Bradlow, E., & Small, D. (2013). New measures of clumpiness for incidence data. *Journal of Applied Statistics*, 40, 2533-2548.
57. Zhang, Y., Bradlow, E., & Small, D. (2014). Predicting Customer Value Using Clumpiness: From RFM to RFMC. *Marketing Science*, 34, 1-14.
58. Ziafat, H., & Shakeri, M. (2014). Using Data Mining Techniques in Customer Segmentation. *Journal of Engineering Research and Applications*, 4, 70-79.
59. Ziph, R. (1971). Psychographics for Market Segmentation. *Journal of Advertising Research*, 11, 3-9.

7 Παράρτημα

Παρατίθενται παρακάτω οι διαδικασίες RFM και RFMV σε γλώσσα προγραμματισμού XML για χρήση στο λογισμικό εξόρυξης δεδομένων RapidMiner Studio.

7.1 Διαδικασία XML RFM

```
<?xml version="1.0" encoding="UTF-8"?><process version="9.10.001">
  <context>
    <input/>
    <output/>
    <macros/>
  </context>
  <operator activated="true" class="process" compatibility="9.10.001" expanded="true"
name="Process">
    <parameter key="logverbosity" value="init"/>
    <parameter key="random_seed" value="2001"/>
    <parameter key="send_mail" value="never"/>
    <parameter key="notification_email" value=""/>
    <parameter key="process_duration_for_mail" value="30"/>
    <parameter key="encoding" value="SYSTEM"/>
    <process expanded="true">
      <operator activated="true" class="retrieve" compatibility="9.10.001" expanded="true"
height="68" name="Retrieve sales" width="90" x="45" y="136">
        <parameter key="repository_entry" value="//Local Repository/sales"/>
      </operator>
      <operator activated="true" class="multiply" compatibility="9.10.001" expanded="true"
height="124" name="Multiply" width="90" x="179" y="85"/>
      <operator activated="true" class="aggregate" compatibility="9.10.001" expanded="true"
height="82" name="Aggregate for F" width="90" x="380" y="136">
        <parameter key="use_default_aggregation" value="false"/>
        <parameter key="attribute_filter_type" value="all"/>
        <parameter key="attribute" value=""/>
        <parameter key="attributes" value=""/>
        <parameter key="use_except_expression" value="false"/>
        <parameter key="value_type" value="attribute_value"/>
        <parameter key="use_value_type_exception" value="false"/>
        <parameter key="except_value_type" value="time"/>
        <parameter key="block_type" value="attribute_block"/>
      </operator>
    </process>
  </operator>
</process>
```

```

<parameter key="use_block_type_exception" value="false"/>
<parameter key="except_block_type" value="value_matrix_row_start"/>
<parameter key="invert_selection" value="false"/>
<parameter key="include_special_attributes" value="false"/>
<parameter key="default_aggregation_function" value="average"/>
<list key="aggregation_attributes">
  <parameter key="date_trns" value="count"/>
</list>
<parameter key="group_by_attributes" value="hh_id"/>
<parameter key="count_all_combinations" value="false"/>
<parameter key="only_distinct" value="false"/>
<parameter key="ignore_missings" value="true"/>
</operator>
<operator activated="true" class="blending:sort" compatibility="9.10.001" expanded="true"
height="82" name="Sort F" width="90" x="581" y="136">
  <list key="sort_by">
    <parameter key="hh_id" value="ascending"/>
  </list>
</operator>
<operator activated="true" class="aggregate" compatibility="9.10.001" expanded="true"
height="82" name="Aggregate for R" width="90" x="380" y="238">
  <parameter key="use_default_aggregation" value="false"/>
  <parameter key="attribute_filter_type" value="all"/>
  <parameter key="attribute" value=""/>
  <parameter key="attributes" value=""/>
  <parameter key="use_except_expression" value="false"/>
  <parameter key="value_type" value="attribute_value"/>
  <parameter key="use_value_type_exception" value="false"/>
  <parameter key="except_value_type" value="time"/>
  <parameter key="block_type" value="attribute_block"/>
  <parameter key="use_block_type_exception" value="false"/>
  <parameter key="except_block_type" value="value_matrix_row_start"/>
  <parameter key="invert_selection" value="false"/>
  <parameter key="include_special_attributes" value="false"/>
  <parameter key="default_aggregation_function" value="average"/>
  <list key="aggregation_attributes">
    <parameter key="date_trns" value="maximum"/>
  </list>
  <parameter key="group_by_attributes" value="hh_id"/>
  <parameter key="count_all_combinations" value="false"/>

```

```

    <parameter key="only_distinct" value="false"/>
    <parameter key="ignore_missings" value="true"/>
  </operator>
  <operator activated="true" class="generate_attributes" compatibility="9.10.001"
expanded="true" height="82" name="Generate Attributes" width="90" x="581" y="238">
    <list key="function_descriptions">
      <parameter key="R_in_months"
value="date_diff([maximum(date_trns)],date_parse_custom("&quot;03.03.2014&quot;,&quot;dd.
MM.yyyy&quot;))/6.048e+8"/>
    </list>
    <parameter key="keep_all" value="true"/>
  </operator>
  <operator activated="true" class="select_attributes" compatibility="9.10.001"
expanded="true" height="82" name="Select Attributes" width="90" x="715" y="238">
    <parameter key="attribute_filter_type" value="subset"/>
    <parameter key="attribute" value=""/>
    <parameter key="attributes" value="hh_id|R_in_months"/>
    <parameter key="use_except_expression" value="false"/>
    <parameter key="value_type" value="attribute_value"/>
    <parameter key="use_value_type_exception" value="false"/>
    <parameter key="except_value_type" value="time"/>
    <parameter key="block_type" value="attribute_block"/>
    <parameter key="use_block_type_exception" value="false"/>
    <parameter key="except_block_type" value="value_matrix_row_start"/>
    <parameter key="invert_selection" value="false"/>
    <parameter key="include_special_attributes" value="false"/>
  </operator>
  <operator activated="true" class="blending:sort" compatibility="9.10.001" expanded="true"
height="82" name="Sort R" width="90" x="849" y="238">
    <list key="sort_by">
      <parameter key="hh_id" value="ascending"/>
    </list>
  </operator>
  <operator activated="true" class="aggregate" compatibility="9.10.001" expanded="true"
height="82" name="Aggregate for M" width="90" x="380" y="34">
    <parameter key="use_default_aggregation" value="false"/>
    <parameter key="attribute_filter_type" value="all"/>
    <parameter key="attribute" value=""/>
    <parameter key="attributes" value=""/>
    <parameter key="use_except_expression" value="false"/>

```

```

<parameter key="value_type" value="attribute_value"/>
<parameter key="use_value_type_exception" value="false"/>
<parameter key="except_value_type" value="time"/>
<parameter key="block_type" value="attribute_block"/>
<parameter key="use_block_type_exception" value="false"/>
<parameter key="except_block_type" value="value_matrix_row_start"/>
<parameter key="invert_selection" value="false"/>
<parameter key="include_special_attributes" value="false"/>
<parameter key="default_aggregation_function" value="average"/>
<list key="aggregation_attributes">
  <parameter key="price" value="sum"/>
</list>
<parameter key="group_by_attributes" value="hh_id"/>
<parameter key="count_all_combinations" value="false"/>
<parameter key="only_distinct" value="false"/>
<parameter key="ignore_missings" value="true"/>
</operator>
<operator activated="true" class="blending:sort" compatibility="9.10.001" expanded="true"
height="82" name="Sort M" width="90" x="581" y="34">
  <list key="sort_by">
    <parameter key="hh_id" value="ascending"/>
  </list>
</operator>
<operator activated="true" class="concurrency:join" compatibility="9.10.001"
expanded="true" height="82" name="Join M & F" width="90" x="782" y="136">
  <parameter key="remove_double_attributes" value="true"/>
  <parameter key="join_type" value="inner"/>
  <parameter key="use_id_attribute_as_key" value="false"/>
  <list key="key_attributes">
    <parameter key="hh_id" value="hh_id"/>
  </list>
  <parameter key="keep_both_join_attributes" value="false"/>
</operator>
<operator activated="true" class="concurrency:join" compatibility="9.10.001"
expanded="true" height="82" name="Join MF & R" width="90" x="983" y="136">
  <parameter key="remove_double_attributes" value="true"/>
  <parameter key="join_type" value="inner"/>
  <parameter key="use_id_attribute_as_key" value="false"/>
  <list key="key_attributes">
    <parameter key="hh_id" value="hh_id"/>
  </list>

```

```

</list>
<parameter key="keep_both_join_attributes" value="false"/>
</operator>
<operator activated="true" class="blending:rename" compatibility="9.10.001"
expanded="true" height="82" name="Rename RFM" width="90" x="1117" y="238">
<list key="rename attributes">
<parameter key="count(date_trns)" value="F"/>
<parameter key="sum(price)" value="M"/>
<parameter key="R_in_months" value="R"/>
</list>
<parameter key="from_attribute" value=""/>
<parameter key="to_attribute" value=""/>
</operator>
<operator activated="true" class="order_attributes" compatibility="9.10.001"
expanded="true" height="82" name="Reorder Attributes" width="90" x="1117" y="34">
<parameter key="sort_mode" value="user specified"/>
<parameter key="attribute_ordering" value="hh_id|R|F|M"/>
<parameter key="use_regular_expressions" value="false"/>
<parameter key="handle_unmatched" value="append"/>
<parameter key="sort_direction" value="ascending"/>
</operator>
<operator activated="true" class="select_attributes" compatibility="9.10.001"
expanded="true" height="82" name="Select Attributes RFM" width="90" x="1251" y="34">
<parameter key="attribute_filter_type" value="subset"/>
<parameter key="attribute" value=""/>
<parameter key="attributes" value="F|M|R"/>
<parameter key="use_except_expression" value="false"/>
<parameter key="value_type" value="attribute_value"/>
<parameter key="use_value_type_exception" value="false"/>
<parameter key="except_value_type" value="time"/>
<parameter key="block_type" value="attribute_block"/>
<parameter key="use_block_type_exception" value="false"/>
<parameter key="except_block_type" value="value_matrix_row_start"/>
<parameter key="invert_selection" value="false"/>
<parameter key="include_special_attributes" value="false"/>
</operator>
<operator activated="true" class="normalize" compatibility="9.10.001" expanded="true"
height="103" name="Normalize" width="90" x="1251" y="136">
<parameter key="return_preprocessing_model" value="false"/>
<parameter key="create_view" value="false"/>

```

```

<parameter key="attribute_filter_type" value="all"/>
<parameter key="attribute" value=""/>
<parameter key="attributes" value=""/>
<parameter key="use_except_expression" value="false"/>
<parameter key="value_type" value="numeric"/>
<parameter key="use_value_type_exception" value="false"/>
<parameter key="except_value_type" value="real"/>
<parameter key="block_type" value="value_series"/>
<parameter key="use_block_type_exception" value="false"/>
<parameter key="except_block_type" value="value_series_end"/>
<parameter key="invert_selection" value="false"/>
<parameter key="include_special_attributes" value="false"/>
<parameter key="method" value="Z-transformation"/>
<parameter key="min" value="0.0"/>
<parameter key="max" value="1.0"/>
<parameter key="allow_negative_values" value="false"/>
</operator>
<operator activated="true" class="anomalydetection:k-NN Global Anomaly Score"
compatibility="3.3.000" expanded="true" height="103" name="k-NN Global Anomaly Score"
width="90" x="1452" y="238">
  <parameter key="k" value="4"/>
  <parameter key="use k-th neighbor distance only (no average)" value="false"/>
  <parameter key="measure_types" value="NumericalMeasures"/>
  <parameter key="mixed_measure" value="MixedEuclideanDistance"/>
  <parameter key="nominal_measure" value="NominalDistance"/>
  <parameter key="numerical_measure" value="EuclideanDistance"/>
  <parameter key="divergence" value="GeneralizedIDivergence"/>
  <parameter key="kernel_type" value="radial"/>
  <parameter key="kernel_gamma" value="1.0"/>
  <parameter key="kernel_sigma1" value="1.0"/>
  <parameter key="kernel_sigma2" value="0.0"/>
  <parameter key="kernel_sigma3" value="2.0"/>
  <parameter key="kernel_degree" value="3.0"/>
  <parameter key="kernel_shift" value="1.0"/>
  <parameter key="kernel_a" value="1.0"/>
  <parameter key="kernel_b" value="0.0"/>
  <parameter key="parallelize evaluation process" value="false"/>
  <parameter key="number of threads" value="4"/>
</operator>

```



```

<operator activated="true" class="filter_examples" compatibility="9.10.001"
expanded="true" height="103" name="Filter Examples" width="90" x="1452" y="85">
  <parameter key="parameter_expression" value=""/>
  <parameter key="condition_class" value="custom_filters"/>
  <parameter key="invert_filter" value="true"/>
  <list key="filters_list">
    <parameter key="filters_entry_key" value="outlier.gt.0\,37743"/>
  </list>
  <parameter key="filters_logic_and" value="true"/>
  <parameter key="filters_check_metadata" value="true"/>
</operator>
<operator activated="true" class="concurrency:k_means" compatibility="9.10.001"
expanded="true" height="82" name="Clustering" width="90" x="1720" y="34">
  <parameter key="add_cluster_attribute" value="true"/>
  <parameter key="add_as_label" value="false"/>
  <parameter key="remove_unlabeled" value="false"/>
  <parameter key="k" value="3"/>
  <parameter key="max_runs" value="30"/>
  <parameter key="determine_good_start_values" value="true"/>
  <parameter key="measure_types" value="NumericalMeasures"/>
  <parameter key="mixed_measure" value="MixedEuclideanDistance"/>
  <parameter key="nominal_measure" value="NominalDistance"/>
  <parameter key="numerical_measure" value="EuclideanDistance"/>
  <parameter key="divergence" value="SquaredEuclideanDistance"/>
  <parameter key="kernel_type" value="radial"/>
  <parameter key="kernel_gamma" value="1.0"/>
  <parameter key="kernel_sigma1" value="1.0"/>
  <parameter key="kernel_sigma2" value="0.0"/>
  <parameter key="kernel_sigma3" value="2.0"/>
  <parameter key="kernel_degree" value="3.0"/>
  <parameter key="kernel_shift" value="1.0"/>
  <parameter key="kernel_a" value="1.0"/>
  <parameter key="kernel_b" value="0.0"/>
  <parameter key="max_optimization_steps" value="100"/>
  <parameter key="use_local_random_seed" value="false"/>
  <parameter key="local_random_seed" value="1992"/>
</operator>
<operator activated="true" class="cluster_distance_performance" compatibility="9.10.001"
expanded="true" height="103" name="Performance" width="90" x="1653" y="187">
  <parameter key="main_criterion" value="Davies Bouldin"/>

```

```

    <parameter key="main_criterion_only" value="false"/>
    <parameter key="normalize" value="false"/>
    <parameter key="maximize" value="false"/>
  </operator>
  <operator      activated="true"      class="model_simulator:cluster_model_visualizer"
compatibility="9.10.001" expanded="true" height="103" name="Cluster Model Visualizer"
width="90" x="1854" y="238"/>
    <connect from_op="Retrieve sales" from_port="output" to_op="Multiply" to_port="input"/>
    <connect from_op="Multiply" from_port="output 1" to_op="Aggregate for M"
to_port="example set input"/>
    <connect from_op="Multiply" from_port="output 2" to_op="Aggregate for F"
to_port="example set input"/>
    <connect from_op="Multiply" from_port="output 3" to_op="Aggregate for R"
to_port="example set input"/>
    <connect from_op="Aggregate for F" from_port="example set output" to_op="Sort F"
to_port="example set input"/>
    <connect from_op="Sort F" from_port="example set output" to_op="Join M & F"
to_port="right"/>
    <connect from_op="Aggregate for R" from_port="example set output" to_op="Generate
Attributes" to_port="example set input"/>
    <connect from_op="Generate Attributes" from_port="example set output" to_op="Select
Attributes" to_port="example set input"/>
    <connect from_op="Select Attributes" from_port="example set output" to_op="Sort R"
to_port="example set input"/>
    <connect from_op="Sort R" from_port="example set output" to_op="Join MF & R"
to_port="right"/>
    <connect from_op="Aggregate for M" from_port="example set output" to_op="Sort M"
to_port="example set input"/>
    <connect from_op="Sort M" from_port="example set output" to_op="Join M & F"
to_port="left"/>
    <connect from_op="Join M & F" from_port="join" to_op="Join MF & R"
to_port="left"/>
    <connect from_op="Join MF & R" from_port="join" to_op="Rename RFM"
to_port="example set input"/>
    <connect from_op="Rename RFM" from_port="example set output" to_op="Reorder
Attributes" to_port="example set input"/>
    <connect from_op="Reorder Attributes" from_port="example set output" to_op="Select
Attributes RFM" to_port="example set input"/>
    <connect from_op="Select Attributes RFM" from_port="example set output"
to_op="Normalize" to_port="example set input"/>

```

```

    <connect from_op="Normalize" from_port="example set output" to_op="k-NN Global
Anomaly Score" to_port="example set"/>
    <connect from_op="k-NN Global Anomaly Score" from_port="example set" to_op="Filter
Examples" to_port="example set input"/>
    <connect from_op="Filter Examples" from_port="example set output" to_op="Clustering"
to_port="example set"/>
    <connect from_op="Filter Examples" from_port="unmatched example set" to_port="result
4"/>
    <connect from_op="Clustering" from_port="cluster model" to_op="Performance"
to_port="cluster model"/>
    <connect from_op="Clustering" from_port="clustered set" to_op="Performance"
to_port="example set"/>
    <connect from_op="Performance" from_port="performance" to_port="result 1"/>
    <connect from_op="Performance" from_port="example set" to_op="Cluster Model
Visualizer" to_port="clustered data"/>
    <connect from_op="Performance" from_port="cluster model" to_op="Cluster Model
Visualizer" to_port="model"/>
    <connect from_op="Cluster Model Visualizer" from_port="visualizer output" to_port="result
2"/>
    <connect from_op="Cluster Model Visualizer" from_port="model output" to_port="result
3"/>
    <portSpacing port="source_input 1" spacing="0"/>
    <portSpacing port="sink_result 1" spacing="0"/>
    <portSpacing port="sink_result 2" spacing="0"/>
    <portSpacing port="sink_result 3" spacing="0"/>
    <portSpacing port="sink_result 4" spacing="0"/>
    <portSpacing port="sink_result 5" spacing="0"/>
  </process>
</operator>
</process>

```

7.2 Διαδικασία XML RFMV

```
<?xml version="1.0" encoding="UTF-8"?><process version="9.10.001">
  <context>
    <input/>
    <output/>
    <macros/>
  </context>
  <operator activated="true" class="process" compatibility="9.10.001" expanded="true"
name="Process">
  <parameter key="logverbosity" value="init"/>
  <parameter key="random_seed" value="2001"/>
  <parameter key="send_mail" value="never"/>
  <parameter key="notification_email" value=""/>
  <parameter key="process_duration_for_mail" value="30"/>
  <parameter key="encoding" value="SYSTEM"/>
  <process expanded="true">
    <operator activated="true" class="retrieve" compatibility="9.10.001" expanded="true"
height="68" name="Retrieve sales" width="90" x="45" y="136">
      <parameter key="repository_entry" value="//Local Repository/sales"/>
    </operator>
    <operator activated="true" class="multiply" compatibility="9.10.001" expanded="true"
height="145" name="Multiply" width="90" x="179" y="85"/>
    <operator activated="true" class="aggregate" compatibility="9.10.001" expanded="true"
height="82" name="Aggregate for F" width="90" x="380" y="136">
      <parameter key="use_default_aggregation" value="false"/>
      <parameter key="attribute_filter_type" value="all"/>
      <parameter key="attribute" value=""/>
      <parameter key="attributes" value=""/>
      <parameter key="use_except_expression" value="false"/>
      <parameter key="value_type" value="attribute_value"/>
      <parameter key="use_value_type_exception" value="false"/>
      <parameter key="except_value_type" value="time"/>
      <parameter key="block_type" value="attribute_block"/>
      <parameter key="use_block_type_exception" value="false"/>
      <parameter key="except_block_type" value="value_matrix_row_start"/>
      <parameter key="invert_selection" value="false"/>
      <parameter key="include_special_attributes" value="false"/>
      <parameter key="default_aggregation_function" value="average"/>
      <list key="aggregation_attributes">
```

```

    <parameter key="date_trns" value="count"/>
  </list>
  <parameter key="group_by_attributes" value="hh_id"/>
  <parameter key="count_all_combinations" value="false"/>
  <parameter key="only_distinct" value="false"/>
  <parameter key="ignore_missings" value="true"/>
</operator>
<operator activated="true" class="blending:sort" compatibility="9.10.001" expanded="true"
height="82" name="Sort F" width="90" x="581" y="136">
  <list key="sort_by">
    <parameter key="hh_id" value="ascending"/>
  </list>
</operator>
<operator activated="true" class="aggregate" compatibility="9.10.001" expanded="true"
height="82" name="Aggregate for R" width="90" x="380" y="238">
  <parameter key="use_default_aggregation" value="false"/>
  <parameter key="attribute_filter_type" value="all"/>
  <parameter key="attribute" value=""/>
  <parameter key="attributes" value=""/>
  <parameter key="use_except_expression" value="false"/>
  <parameter key="value_type" value="attribute_value"/>
  <parameter key="use_value_type_exception" value="false"/>
  <parameter key="except_value_type" value="time"/>
  <parameter key="block_type" value="attribute_block"/>
  <parameter key="use_block_type_exception" value="false"/>
  <parameter key="except_block_type" value="value_matrix_row_start"/>
  <parameter key="invert_selection" value="false"/>
  <parameter key="include_special_attributes" value="false"/>
  <parameter key="default_aggregation_function" value="average"/>
  <list key="aggregation_attributes">
    <parameter key="date_trns" value="maximum"/>
  </list>
  <parameter key="group_by_attributes" value="hh_id"/>
  <parameter key="count_all_combinations" value="false"/>
  <parameter key="only_distinct" value="false"/>
  <parameter key="ignore_missings" value="true"/>
</operator>
<operator activated="true" class="generate_attributes" compatibility="9.10.001"
expanded="true" height="82" name="Generate Attributes" width="90" x="581" y="238">
  <list key="function_descriptions">

```

```

        <parameter key="R_in_months"
value="date_diff([maximum(date_trns)],date_parse_custom("&quot;03.03.2014&quot;,&quot;dd.
MM.yyyy&quot;))/6.048e+8"/>
    </list>
    <parameter key="keep_all" value="true"/>
</operator>
<operator activated="true" class="select_attributes" compatibility="9.10.001"
expanded="true" height="82" name="Select Attributes" width="90" x="715" y="238">
    <parameter key="attribute_filter_type" value="subset"/>
    <parameter key="attribute" value=""/>
    <parameter key="attributes" value="hh_id|R_in_months"/>
    <parameter key="use_except_expression" value="false"/>
    <parameter key="value_type" value="attribute_value"/>
    <parameter key="use_value_type_exception" value="false"/>
    <parameter key="except_value_type" value="time"/>
    <parameter key="block_type" value="attribute_block"/>
    <parameter key="use_block_type_exception" value="false"/>
    <parameter key="except_block_type" value="value_matrix_row_start"/>
    <parameter key="invert_selection" value="false"/>
    <parameter key="include_special_attributes" value="false"/>
</operator>
<operator activated="true" class="blending:sort" compatibility="9.10.001" expanded="true"
height="82" name="Sort R" width="90" x="849" y="238">
    <list key="sort_by">
        <parameter key="hh_id" value="ascending"/>
    </list>
</operator>
<operator activated="true" class="aggregate" compatibility="9.10.001" expanded="true"
height="82" name="Aggregate for M" width="90" x="380" y="34">
    <parameter key="use_default_aggregation" value="false"/>
    <parameter key="attribute_filter_type" value="all"/>
    <parameter key="attribute" value=""/>
    <parameter key="attributes" value=""/>
    <parameter key="use_except_expression" value="false"/>
    <parameter key="value_type" value="attribute_value"/>
    <parameter key="use_value_type_exception" value="false"/>
    <parameter key="except_value_type" value="time"/>
    <parameter key="block_type" value="attribute_block"/>
    <parameter key="use_block_type_exception" value="false"/>
    <parameter key="except_block_type" value="value_matrix_row_start"/>

```

```

<parameter key="invert_selection" value="false"/>
<parameter key="include_special_attributes" value="false"/>
<parameter key="default_aggregation_function" value="average"/>
<list key="aggregation_attributes">
  <parameter key="price" value="sum"/>
</list>
<parameter key="group_by_attributes" value="hh_id"/>
<parameter key="count_all_combinations" value="false"/>
<parameter key="only_distinct" value="false"/>
<parameter key="ignore_missings" value="true"/>
</operator>
<operator activated="true" class="blending:sort" compatibility="9.10.001" expanded="true"
height="82" name="Sort M" width="90" x="581" y="34">
  <list key="sort_by">
    <parameter key="hh_id" value="ascending"/>
  </list>
</operator>
<operator activated="true" class="concurrency:join" compatibility="9.10.001"
expanded="true" height="82" name="Join M & F" width="90" x="782" y="136">
  <parameter key="remove_double_attributes" value="true"/>
  <parameter key="join_type" value="inner"/>
  <parameter key="use_id_attribute_as_key" value="false"/>
  <list key="key_attributes">
    <parameter key="hh_id" value="hh_id"/>
  </list>
  <parameter key="keep_both_join_attributes" value="false"/>
</operator>
<operator activated="true" class="concurrency:join" compatibility="9.10.001"
expanded="true" height="82" name="Join MF & R" width="90" x="983" y="136">
  <parameter key="remove_double_attributes" value="true"/>
  <parameter key="join_type" value="inner"/>
  <parameter key="use_id_attribute_as_key" value="false"/>
  <list key="key_attributes">
    <parameter key="hh_id" value="hh_id"/>
  </list>
  <parameter key="keep_both_join_attributes" value="false"/>
</operator>
<operator activated="true" class="retrieve" compatibility="9.10.001" expanded="true"
height="68" name="Retrieve" width="90" x="45" y="442">
  <parameter key="repository_entry" value="products"/>

```

```

</operator>
<operator activated="true" class="numerical_to_polynomial" compatibility="9.10.001"
expanded="true" height="82" name="Numerical to Polynomial" width="90" x="179" y="442">
  <parameter key="attribute_filter_type" value="subset"/>
  <parameter key="attribute" value=""/>
  <parameter key="attributes" value="item_class_cd|item_sc_cd|item_id"/>
  <parameter key="use_except_expression" value="false"/>
  <parameter key="value_type" value="numeric"/>
  <parameter key="use_value_type_exception" value="false"/>
  <parameter key="except_value_type" value="real"/>
  <parameter key="block_type" value="value_series"/>
  <parameter key="use_block_type_exception" value="false"/>
  <parameter key="except_block_type" value="value_series_end"/>
  <parameter key="invert_selection" value="false"/>
  <parameter key="include_special_attributes" value="false"/>
</operator>
<operator activated="true" class="generate_attributes" compatibility="9.10.001"
expanded="true" height="82" name="Generate Attributes (Product Category)" width="90"
x="313" y="442">
  <list key="function_descriptions">
    <parameter key="product_category"
value="concat(item_grp_cd,&quot;;&quot;;item_class_cd,&quot;;&quot;;item_sc_cd)"/>
  </list>
  <parameter key="keep_all" value="true"/>
</operator>
<operator activated="true" class="numerical_to_polynomial" compatibility="9.10.001"
expanded="true" height="82" name="Numerical to Polynomial ITEM_ID" width="90" x="313"
y="340">
  <parameter key="attribute_filter_type" value="single"/>
  <parameter key="attribute" value="item_id"/>
  <parameter key="attributes" value=""/>
  <parameter key="use_except_expression" value="false"/>
  <parameter key="value_type" value="numeric"/>
  <parameter key="use_value_type_exception" value="false"/>
  <parameter key="except_value_type" value="real"/>
  <parameter key="block_type" value="value_series"/>
  <parameter key="use_block_type_exception" value="false"/>
  <parameter key="except_block_type" value="value_series_end"/>
  <parameter key="invert_selection" value="false"/>
  <parameter key="include_special_attributes" value="false"/>

```



```

</operator>
<operator activated="true" class="concurrency:join" compatibility="9.10.001"
expanded="true" height="82" name="Join Sales & Products" width="90" x="514" y="340">
  <parameter key="remove_double_attributes" value="true"/>
  <parameter key="join_type" value="inner"/>
  <parameter key="use_id_attribute_as_key" value="false"/>
  <list key="key_attributes">
    <parameter key="item_id" value="item_id"/>
  </list>
  <parameter key="keep_both_join_attributes" value="false"/>
</operator>
<operator activated="true" class="select_attributes" compatibility="9.10.001"
expanded="true" height="82" name="Select Attributes (V, Date, ID)" width="90" x="648"
y="340">
  <parameter key="attribute_filter_type" value="subset"/>
  <parameter key="attribute" value=""/>
  <parameter key="attributes" value="hh_id|product_category|date_trns"/>
  <parameter key="use_except_expression" value="false"/>
  <parameter key="value_type" value="attribute_value"/>
  <parameter key="use_value_type_exception" value="false"/>
  <parameter key="except_value_type" value="time"/>
  <parameter key="block_type" value="attribute_block"/>
  <parameter key="use_block_type_exception" value="false"/>
  <parameter key="except_block_type" value="value_matrix_row_start"/>
  <parameter key="invert_selection" value="false"/>
  <parameter key="include_special_attributes" value="false"/>
</operator>
<operator activated="true" class="aggregate" compatibility="9.10.001" expanded="true"
height="82" name="Aggregate for V" width="90" x="849" y="340">
  <parameter key="use_default_aggregation" value="false"/>
  <parameter key="attribute_filter_type" value="all"/>
  <parameter key="attribute" value=""/>
  <parameter key="attributes" value=""/>
  <parameter key="use_except_expression" value="false"/>
  <parameter key="value_type" value="attribute_value"/>
  <parameter key="use_value_type_exception" value="false"/>
  <parameter key="except_value_type" value="time"/>
  <parameter key="block_type" value="attribute_block"/>
  <parameter key="use_block_type_exception" value="false"/>
  <parameter key="except_block_type" value="value_matrix_row_start"/>

```

```

<parameter key="invert_selection" value="false"/>
<parameter key="include_special_attributes" value="false"/>
<parameter key="default_aggregation_function" value="average"/>
<list key="aggregation_attributes">
  <parameter key="product_category" value="count"/>
</list>
<parameter key="group_by_attributes" value="hh_id"/>
<parameter key="count_all_combinations" value="false"/>
<parameter key="only_distinct" value="true"/>
<parameter key="ignore_missings" value="true"/>
</operator>
<operator activated="true" class="concurrency:join" compatibility="9.10.001"
expanded="true" height="82" name="Join RFM & V" width="90" x="1050" y="340">
  <parameter key="remove_double_attributes" value="true"/>
  <parameter key="join_type" value="inner"/>
  <parameter key="use_id_attribute_as_key" value="false"/>
  <list key="key_attributes">
    <parameter key="hh_id" value="hh_id"/>
  </list>
  <parameter key="keep_both_join_attributes" value="false"/>
</operator>
<operator activated="true" class="blending:rename" compatibility="9.10.001"
expanded="true" height="82" name="Rename" width="90" x="1117" y="136">
  <list key="rename_attributes">
    <parameter key="count(date_trns)" value="F"/>
    <parameter key="sum(price)" value="M"/>
    <parameter key="R_in_months" value="R"/>
    <parameter key="count(product_category)" value="V"/>
  </list>
  <parameter key="from_attribute" value=""/>
  <parameter key="to_attribute" value=""/>
</operator>
<operator activated="true" class="order_attributes" compatibility="9.10.001"
expanded="true" height="82" name="Reorder Attributes" width="90" x="1117" y="34">
  <parameter key="sort_mode" value="user specified"/>
  <parameter key="attribute_ordering" value="hh_id|R|F|M|V"/>
  <parameter key="use_regular_expressions" value="false"/>
  <parameter key="handle_unmatched" value="append"/>
  <parameter key="sort_direction" value="ascending"/>
</operator>

```

```

<operator activated="true" class="select_attributes" compatibility="9.10.001"
expanded="true" height="82" name="Select Attributes (RFMV)" width="90" x="1251" y="34">
  <parameter key="attribute_filter_type" value="subset"/>
  <parameter key="attribute" value=""/>
  <parameter key="attributes" value="F|M|R|V"/>
  <parameter key="use_except_expression" value="false"/>
  <parameter key="value_type" value="attribute_value"/>
  <parameter key="use_value_type_exception" value="false"/>
  <parameter key="except_value_type" value="time"/>
  <parameter key="block_type" value="attribute_block"/>
  <parameter key="use_block_type_exception" value="false"/>
  <parameter key="except_block_type" value="value_matrix_row_start"/>
  <parameter key="invert_selection" value="false"/>
  <parameter key="include_special_attributes" value="false"/>
</operator>
<operator activated="true" class="normalize" compatibility="9.10.001" expanded="true"
height="103" name="Normalize" width="90" x="1251" y="136">
  <parameter key="return_preprocessing_model" value="false"/>
  <parameter key="create_view" value="false"/>
  <parameter key="attribute_filter_type" value="all"/>
  <parameter key="attribute" value=""/>
  <parameter key="attributes" value=""/>
  <parameter key="use_except_expression" value="false"/>
  <parameter key="value_type" value="numeric"/>
  <parameter key="use_value_type_exception" value="false"/>
  <parameter key="except_value_type" value="real"/>
  <parameter key="block_type" value="value_series"/>
  <parameter key="use_block_type_exception" value="false"/>
  <parameter key="except_block_type" value="value_series_end"/>
  <parameter key="invert_selection" value="false"/>
  <parameter key="include_special_attributes" value="false"/>
  <parameter key="method" value="Z-transformation"/>
  <parameter key="min" value="0.0"/>
  <parameter key="max" value="1.0"/>
  <parameter key="allow_negative_values" value="false"/>
</operator>
<operator activated="true" class="anomalydetection:k-NN Global Anomaly Score"
compatibility="3.3.000" expanded="true" height="103" name="k-NN Global Anomaly Score"
width="90" x="1452" y="238">
  <parameter key="k" value="4"/>

```

```

<parameter key="use k-th neighbor distance only (no average)" value="false"/>
<parameter key="measure_types" value="NumericalMeasures"/>
<parameter key="mixed_measure" value="MixedEuclideanDistance"/>
<parameter key="nominal_measure" value="NominalDistance"/>
<parameter key="numerical_measure" value="EuclideanDistance"/>
<parameter key="divergence" value="GeneralizedIDivergence"/>
<parameter key="kernel_type" value="radial"/>
<parameter key="kernel_gamma" value="1.0"/>
<parameter key="kernel_sigma1" value="1.0"/>
<parameter key="kernel_sigma2" value="0.0"/>
<parameter key="kernel_sigma3" value="2.0"/>
<parameter key="kernel_degree" value="3.0"/>
<parameter key="kernel_shift" value="1.0"/>
<parameter key="kernel_a" value="1.0"/>
<parameter key="kernel_b" value="0.0"/>
<parameter key="parallelize evaluation process" value="false"/>
<parameter key="number of threads" value="4"/>
</operator>
<operator activated="true" class="filter_examples" compatibility="9.10.001"
expanded="true" height="103" name="Filter Examples" width="90" x="1452" y="85">
  <parameter key="parameter_expression" value=""/>
  <parameter key="condition_class" value="custom_filters"/>
  <parameter key="invert_filter" value="true"/>
  <list key="filters_list">
    <parameter key="filters_entry_key" value="outlier.gt.0\,53"/>
  </list>
  <parameter key="filters_logic_and" value="true"/>
  <parameter key="filters_check_metadata" value="true"/>
</operator>
<operator activated="true" class="concurrency:k_means" compatibility="9.10.001"
expanded="true" height="82" name="Clustering k-Means" width="90" x="1720" y="34">
  <parameter key="add_cluster_attribute" value="true"/>
  <parameter key="add_as_label" value="false"/>
  <parameter key="remove_unlabeled" value="false"/>
  <parameter key="k" value="3"/>
  <parameter key="max_runs" value="30"/>
  <parameter key="determine_good_start_values" value="true"/>
  <parameter key="measure_types" value="NumericalMeasures"/>
  <parameter key="mixed_measure" value="MixedEuclideanDistance"/>
  <parameter key="nominal_measure" value="NominalDistance"/>

```

```

<parameter key="numerical_measure" value="EuclideanDistance"/>
<parameter key="divergence" value="SquaredEuclideanDistance"/>
<parameter key="kernel_type" value="radial"/>
<parameter key="kernel_gamma" value="1.0"/>
<parameter key="kernel_sigma1" value="1.0"/>
<parameter key="kernel_sigma2" value="0.0"/>
<parameter key="kernel_sigma3" value="2.0"/>
<parameter key="kernel_degree" value="3.0"/>
<parameter key="kernel_shift" value="1.0"/>
<parameter key="kernel_a" value="1.0"/>
<parameter key="kernel_b" value="0.0"/>
<parameter key="max_optimization_steps" value="100"/>
<parameter key="use_local_random_seed" value="false"/>
<parameter key="local_random_seed" value="1992"/>
</operator>
<operator activated="true" class="cluster_distance_performance" compatibility="9.10.001"
expanded="true" height="103" name="Performance" width="90" x="1653" y="187">
  <parameter key="main_criterion" value="Davies Bouldin"/>
  <parameter key="main_criterion_only" value="false"/>
  <parameter key="normalize" value="false"/>
  <parameter key="maximize" value="false"/>
</operator>
<operator activated="true" class="model_simulator:cluster_model_visualizer"
compatibility="9.10.001" expanded="true" height="103" name="Cluster Model Visualizer"
width="90" x="1854" y="238"/>
  <connect from_op="Retrieve sales" from_port="output" to_op="Multiply" to_port="input"/>
  <connect from_op="Multiply" from_port="output 1" to_op="Aggregate for M"
to_port="example set input"/>
  <connect from_op="Multiply" from_port="output 2" to_op="Aggregate for F"
to_port="example set input"/>
  <connect from_op="Multiply" from_port="output 3" to_op="Aggregate for R"
to_port="example set input"/>
  <connect from_op="Multiply" from_port="output 4" to_op="Numerical to Polynomial
ITEM_ID" to_port="example set input"/>
  <connect from_op="Aggregate for F" from_port="example set output" to_op="Sort F"
to_port="example set input"/>
  <connect from_op="Sort F" from_port="example set output" to_op="Join M & F"
to_port="right"/>
  <connect from_op="Aggregate for R" from_port="example set output" to_op="Generate
Attributes" to_port="example set input"/>

```

```

    <connect from_op="Generate Attributes" from_port="example set output" to_op="Select
Attributes" to_port="example set input"/>
    <connect from_op="Select Attributes" from_port="example set output" to_op="Sort R"
to_port="example set input"/>
    <connect from_op="Sort R" from_port="example set output" to_op="Join MF & R"
to_port="right"/>
    <connect from_op="Aggregate for M" from_port="example set output" to_op="Sort M"
to_port="example set input"/>
    <connect from_op="Sort M" from_port="example set output" to_op="Join M & F"
to_port="left"/>
    <connect from_op="Join M & F" from_port="join" to_op="Join MF & R"
to_port="left"/>
    <connect from_op="Join MF & R" from_port="join" to_op="Join RFM & V"
to_port="left"/>
    <connect from_op="Retrieve" from_port="output" to_op="Numerical to Polynominal"
to_port="example set input"/>
    <connect from_op="Numerical to Polynominal" from_port="example set output"
to_op="Generate Attributes (Product Category)" to_port="example set input"/>
    <connect from_op="Generate Attributes (Product Category)" from_port="example set output"
to_op="Join Sales & Products" to_port="right"/>
    <connect from_op="Numerical to Polynominal ITEM_ID" from_port="example set output"
to_op="Join Sales & Products" to_port="left"/>
    <connect from_op="Join Sales & Products" from_port="join" to_op="Select Attributes
(V, Date, ID)" to_port="example set input"/>
    <connect from_op="Select Attributes (V, Date, ID)" from_port="example set output"
to_op="Aggregate for V" to_port="example set input"/>
    <connect from_op="Aggregate for V" from_port="example set output" to_op="Join RFM
& V" to_port="right"/>
    <connect from_op="Join RFM & V" from_port="join" to_op="Rename"
to_port="example set input"/>
    <connect from_op="Rename" from_port="example set output" to_op="Reorder Attributes"
to_port="example set input"/>
    <connect from_op="Reorder Attributes" from_port="example set output" to_op="Select
Attributes (RFMV)" to_port="example set input"/>
    <connect from_op="Select Attributes (RFMV)" from_port="example set output"
to_op="Normalize" to_port="example set input"/>
    <connect from_op="Normalize" from_port="example set output" to_op="k-NN Global
Anomaly Score" to_port="example set"/>
    <connect from_op="k-NN Global Anomaly Score" from_port="example set" to_op="Filter
Examples" to_port="example set input"/>

```

```

    <connect from_op="Filter Examples" from_port="example set output" to_op="Clustering k-
Means" to_port="example set"/>
    <connect from_op="Filter Examples" from_port="unmatched example set" to_op="result
4"/>
    <connect from_op="Clustering k-Means" from_port="cluster model" to_op="Performance"
to_port="cluster model"/>
    <connect from_op="Clustering k-Means" from_port="clustered set" to_op="Performance"
to_port="example set"/>
    <connect from_op="Performance" from_port="performance" to_port="result 1"/>
    <connect from_op="Performance" from_port="example set" to_op="Cluster Model
Visualizer" to_port="clustered data"/>
    <connect from_op="Performance" from_port="cluster model" to_op="Cluster Model
Visualizer" to_port="model"/>
    <connect from_op="Cluster Model Visualizer" from_port="visualizer output" to_port="result
2"/>
    <connect from_op="Cluster Model Visualizer" from_port="model output" to_port="result
3"/>
    <portSpacing port="source_input 1" spacing="0"/>
    <portSpacing port="sink_result 1" spacing="0"/>
    <portSpacing port="sink_result 2" spacing="0"/>
    <portSpacing port="sink_result 3" spacing="0"/>
    <portSpacing port="sink_result 4" spacing="0"/>
    <portSpacing port="sink_result 5" spacing="0"/>
</process>
</operator>
</process>

```