



Πολυτεχνείο Κρήτης

Σχολή Μηχανικών Παραγωγής και Διοίκησης

Ανάλυση συμπεριφοράς καταναλωτή με χρήση μεθόδων πολυκριτήριας ανάλυσης και μηχανικής μάθησης

Διπλωματική Εργασία

Κωνσταντίνα Καδίτη

Επιβλέπων

Ματσατσίνης Νικόλαος, Καθηγητής

Χανιά, Σεπτέμβριος 2022

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον Καθηγητή μου κ. Ματσατσίνη Νικόλαο για την ευκαιρία που μου έδωσε να πραγματοποιήσω την εκπόνηση της διπλωματικής μου εργασίας σε ένα αντικείμενο που μου αρέσει πολύ. Χωρίς την πολύτιμη καθοδήγηση και τις συμβουλές του, θα ήταν ακατόρθωτο το τελικό αποτέλεσμα. Επίσης, δεν θα μπορούσα να παραλείψω να ευχαριστήσω την υποψήφια Διδάκτορα και φίλη Καλαφάτη Φωτεινή, η οποία με στήριξε σε κάθε βήμα της διπλωματικής εργασίας, καθώς και κατά τη διάρκεια των προηγούμενων χρόνων, με κατανόηση και ενθάρρυνση. Τέλος, όφειλω ένα μεγάλο ευχαριστώ στην οικογένειά μου, στους φίλους μου και στον σύντροφό μου Σταύρο Λελούδα για την συναισθηματική και οικονομική στήριξη και ασφάλεια που μου παρείχαν όλα αυτά τα χρόνια.

Περίληψη

Στον αιώνα που διανύουμε τα δεδομένα αντιστοιχούν σε πηγή πλούτου. Ο λόγος που αυτό συμβαίνει, είναι η ραγδαία εξέλιξη της τεχνολογίας και ειδικότερα του διαδικτύου. Τα τελευταία χρόνια λόγω της εξάπλωσης της ασθένειας Covid-19, όλο και περισσότεροι χρησιμοποίησαν το διαδίκτυο για αγορές, ακόμα και βασικών ειδών πρώτης ανάγκης, κατα την διάρκεια των lockdown που επιβλήθηκαν παγκοσμίως. Καθεμία συναλλαγή εμπεριέχει δεδομένα για το πρόσωπο που την έκανε, όπως π.χ. φύλο, ηλικία, μέρος καθώς και τα προϊόντα της ίδιας της αγοράς του. Ωστόσο, αυτά τα δεδομένα δεν μπορούν να αξιοποιηθούν αυτούσια χωρίς επεξεργασία. Η επιστήμη της ανάλυσης δεδομένων επικεντρώνεται σε αυτό ακριβώς. Στο πώς, δηλαδή, θα μετατρέψει τα δεδομένα σε πληροφορία αξιοποιώντας τις δυνατότητες της στατιστικής, της εξόρυξης δεδομένων, της πολυκριτήριας ανάλυσης αποφάσεων, της οπτικοποίησης των αποτελεσμάτων, προκειμένου να βοηθηθούν οι εταιρίες να καταφέρουν να μελετήσουν τον ανταγωνισμό, να διαμορφώσουν τις πολιτικές μάρκετινγκ τους, να αυξήσουν τις πωλήσεις τους μέσω πιο στοχευμένων διαφημίσεων, βελτιστοποίησης των ήδη υπάρχοντων προϊόντων ή μέσω δημιουργίας νέων προϊόντων που θα καλύψουν κάποιο κενό στην αγορά.

Στόχος της εργασίας αυτής είναι η ανάλυση συμπεριφοράς καταναλωτών μέσω μεθόδων μηχανικής μάθησης και πολυκριτήριας ανάλυσης. Για την συλλογή δεδομένων θα χρησιμοποιηθούν ερωτηματολόγια που θα συμπληρωθούν από καταναλωτές super markets στην περιοχή των Χανίων. Τα δεδομένα θα συλλεχθούν με ερωτηματολόγια που θα διανεμηθούν σε καταναλωτές δια ζώσης και από τα μέσα κοινωνικής δικτύωσης. Για την επεξεργασία των δεδομένων θα χρησιμοποιηθούν αλγόριθμοι συσταδοποίησης όπως ο k-Means, καθώς και μέθοδοι ανάλυσης συσχετίσεων για τη διενέργεια Market Basket Analysis. Το Market Basket Analysis είναι μέθοδος εξόρυξης δεδομένων και χρησιμοποιείται ώστε να ευρεθούν μοτίβα και συσχετίσεις με τα οποία οι καταναλωτές αγοράζουν. Το μοτίβο μπορεί να αναλυθεί χρησιμοποιώντας το ιστορικό αγορών των καταναλωτών ή την ομαδοποίηση των προϊόντων που συνήθως αγοράζονται μαζί. Γι' αυτό το λόγο, χρησιμοποιούνται αλγόριθμοι συσχέτισης, όπως ο FP Growth. Θα χρησιμοποιηθούν επίσης αλγόριθμοι πολυκριτήριας ανάλυσης, όπως η Utastar για την ανάλυση συμπεριφοράς καταναλωτών με στόχο την μελέτη του ανταγωνισμού.

Περιεχόμενα

Ευχαριστίες.....	3
Περίληψη.....	4
Κατάλογος Πινάκων	7
Κατάλογος Γραφημάτων	8
Κεφάλαιο 1 ΕΙΣΑΓΩΓΗ	10
1.1 ΣΚΟΠΟΣ ΤΗΣ ΕΡΓΑΣΙΑΣ	10
1.2 ΔΟΜΗ ΤΗΣ ΕΡΓΑΣΙΑΣ	10
Κεφάλαιο 2 ΥΦΙΣΤΑΜΕΝΗ ΚΑΤΑΣΤΑΣΗ.....	11
2.1 ΕΙΣΑΓΩΓΗ	11
2.2 ΥΦΙΣΤΑΜΕΝΗ ΚΑΤΑΣΤΑΣΗ	11
2.3 ΚΑΘΟΡΙΣΜΟΣ ΤΟΥ ΠΡΟΒΛΗΜΑΤΟΣ.....	12
Κεφάλαιο 3 ΜΕΘΟΔΟΛΟΓΙΑ	13
3.1 ΕΙΣΑΓΩΓΗ	13
3.2 ΔΗΜΙΟΥΡΓΙΑ ΒΑΣΗΣ ΔΕΔΟΜΕΝΩΝ	14
3.3 ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ	14
3.4 ΠΡΟΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ	15
3.4.1 Καθαρισμός Δεδομένων	15
3.4.2 Μετασχηματισμός και Διακριτοποίηση Δεδομένων.....	17
3.4.3 Μείωση Δεδομένων.....	17
3.5 ΠΟΛΥΚΡΙΤΗΡΙΑ ΑΝΑΛΥΣΗ UTASTAR.....	17
3.5.1 Εισαγωγή.....	17
3.5.2 Μέθοδος UTASTAR	17
3.6 ΣΥΣΤΑΔΟΠΟΙΗΣΗ	20
3.6.1 Εισαγωγή.....	20
3.6.2 Αλγόριθμος K-means	21
3.7 MARKET BASKET ANALYSIS	22
3.7.1 Εισαγωγή.....	22
3.7.2 Market Basket Analysis	22
Κεφάλαιο 4 ΑΠΟΤΕΛΕΣΜΑΤΑ	25

4.1.	ΕΙΣΑΓΩΓΗ	25
4.2.	ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ	25
4.3.	ΠΡΟΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ	33
4.3.1.	Εισαγωγή δεδομένων στο RapidMiner.....	33
4.3.2.	Προεπεξεργασία Δεδομένων	37
4.4.	ΠΟΛΥΚΡΙΤΗΡΙΑ ΑΝΑΛΥΣΗ UTASTAR.....	44
4.4.1.	Εισαγωγή.....	44
4.4.2.	Εφαρμογή UTASTAR.....	45
4.5.	ΣΥΣΤΑΔΟΠΟΙΗΣΗ	51
4.5.1.	Συσταδοποίηση με χρήση αλγορίθμων MARKEK.....	51
4.5.2.	Αλγόριθμος Συσταδοποίησης K-means.....	55
4.6.	MARKET BASKET ANALYSIS	61
4.6.1.	Εισαγωγή.....	61
4.6.2.	Αποτελέσματα Market Basket Analysis για το Super Market «ΑΒ Βασιλόπουλος»	62
4.6.3.	Αποτελέσματα Market Basket Analysis για το Super Market «Σκλαβενίτης».....	65
4.6.4.	Αποτελέσματα Market Basket Analysis για το Super Market «Χαλκιαδάκης»	67
4.6.5.	Αποτελέσματα Market Basket Analysis για το Super Market «LIDL» ..	70
4.6.6.	Αποτελέσματα Market Basket Analysis για το Super Market «SYNKA»	72
4.6.7.	Αποτελέσματα Market Basket Analysis για την επιλογή «Κανένα».....	75
	Κεφάλαιο 5 ΣΥΜΠΕΡΑΣΜΑΤΑ – ΠΡΟΤΑΣΕΙΣ	79
	ΔΙΕΘΝΗΣ ΒΙΒΛΙΟΓΡΑΦΙΑ.....	82
	ΕΛΛΗΝΙΚΗ ΒΙΒΛΙΟΓΡΑΦΙΑ	83
	ΠΑΡΑΡΤΗΜΑ.....	84
	Ερωτηματολόγιο	84

Κατάλογος Πινάκων

Πίνακας 4-1: Διορθωμένες απαντήσεις	40
Πίνακας 4-2: Μέσα βάρη κριτηρίων	45
Πίνακας 4-3: Μέση ολική χρησιμότητα εναλλακτικών	46
Πίνακας 4-4: Συστάδες καταναλωτών με βάση τα κριτήρια επιλογής super market ...	52
Πίνακας 4-5: Δείκτης Silhouette για το αρχείο “Utilities”	56
Πίνακας 4-6: Δείκτης Silhouette για το αρχείο “Weights”	56
Πίνακας 4-7: Παρουσίαση των 10 πρώτων πιο συχνά εμφανιζόμενων προϊόντων	63
Πίνακας 4-8: Παρουσίαση των κανόνων συσχέτισης	63
Πίνακας 4-9: Παρουσίαση των 10 πρώτων πιο συχνά εμφανιζόμενων προϊόντων	65
Πίνακας 4-10: Παρουσίαση των κανόνων συσχέτισης	66
Πίνακας 4-11: Παρουσίαση των 10 πρώτων πιο συχνά εμφανιζόμενων προϊόντων ...	68
Πίνακας 4-12: Παρουσίαση των κανόνων συσχέτισης	69
Πίνακας 4-13: Παρουσίαση των 10 πρώτων πιο συχνά εμφανιζόμενων προϊόντων ...	70
Πίνακας 4-14: Παρουσίαση κανόνων συσχέτισης	71
Πίνακας 4-16: Παρουσίαση των πιο συχνά εμφανιζόμενων προϊόντων	73
Πίνακας 4-17: Παρουσίαση κανόνων συσχέτισης	73
Πίνακας 4-18: Παρουσίαση των πιο συχνά εμφανιζόμενων προϊόντων	75
Πίνακας 4-19: Παρουσίαση κανόνων συσχέτισης	76
Πίνακας 5-1: Συνοπτική παρουσίαση κανόνων συσχέτισης	80

Κατάλογος Γραφημάτων

Γράφημα 4-1: Ποσοστιαία κατανομή φύλων.....	25
Γράφημα 4-2: Ηλικιακή κατανομή γυναικών.....	26
Γράφημα 4-3: Ηλικιακή κατανομή ανδρών.....	26
Γράφημα 4-4: Μορφωτικό επίπεδο γυναικών.....	27
Γράφημα 4-5: Μορφωτικό επίπεδο ανδρών.....	27
Γράφημα 4-6: Επάγγελμα Γυναικών	28
Γράφημα 4-7: Επάγγελμα Ανδρών	28
Γράφημα 4-8: Ετήσιο Οικογενειακό Εισόδημα	29
Γράφημα 4-9: Μέλη νοικοκυριού για τα οποία γίνονται αγορές από το super market	29
Γράφημα 4-10: Συχνότητα αγορών από super market	30
Γράφημα 4-11: Μέση μηνιαία δαπάνη καταναλωτών που επισκέπτονται το super market 1-2 φορές την εβδομάδα	30
Γράφημα 4-12: Μέση μηνιαία δαπάνη καταναλωτών που επισκέπτονται το super market 1-3 φορές το μήνα.....	31
Γράφημα 4-13: Μέση μηνιαία δαπάνη καταναλωτών που επισκέπτονται το super market 3-5 φορές την εβδομάδα	31
Γράφημα 4-14: Μέση μηνιαία δαπάνη καταναλωτών που επισκέπτονται το super market καθημερινά	32
Γράφημα 4-15: Μέση ολική χρησιμότητα εναλλακτικών κατοίκων Ακρωτηρίου	47
Γράφημα 4-16: Μέση ολική χρησιμότητα εναλλακτικών κατοίκων Δικαστηρίων	47
Γράφημα 4-17: Μέση ολική χρησιμότητα εναλλακτικών κατοίκων Δαράτσου.....	48
Γράφημα 4-18: Μέση ολική χρησιμότητα εναλλακτικών κατοίκων Κέντρου	48
Γράφημα 4-19: Μέση ολική χρησιμότητα εναλλακτικών κατοίκων Νέας Χώρας	49
Γράφημα 4-20: Μέση ολική χρησιμότητα εναλλακτικών κατοίκων Λενταριανών.....	49
Γράφημα 4-21: Μέση ολική χρησιμότητα εναλλακτικών κατοίκων Σούδας	50
Γράφημα 4-22: Μέση ολική χρησιμότητα εναλλακτικών κατοίκων Χαλέπας	50
Γράφημα 4-23: Βάρη κριτηρίων- Συστάδα 1.....	57
Γράφημα 4-24: Βάρη κριτηρίων-Συστάδα 2.....	58
Γράφημα 4-25: Βάρη κριτηρίων-Συστάδα 3.....	58
Γράφημα 4-26: Ολικές χρησιμότητες εναλλακτικών επιλογών-Συστάδα 1.....	59
Γράφημα 4-27: Ολικές χρησιμότητες εναλλακτικών επιλογών-Συστάδα 2.....	60
Γράφημα 4-28: Ολικές χρησιμότητες εναλλακτικών επιλογών-Συστάδα 3.....	60

Κατάλογος Εικόνων

Εικόνα 4-1: Αρχική οθόνη RapidMiner	33
Εικόνα 4-2: Αρχική οθόνη	34
Εικόνα 4-3: Δημιουργία Αποθετηρίου	34
Εικόνα 4-4: Εισαγωγή υποφακέλων στο αποθετήριο	35
Εικόνα 4-5: Εισαγωγή Δεδομένων (Βήμα 1 ^ο)	36
Εικόνα 4-6: Επεξεργασία στηλών πριν την αποθήκευση	36
Εικόνα 4-7: Αποθήκευση δεδομένων	37
Εικόνα 4-8: Operator "Read Excel"	37
Εικόνα 4-9: Εμφάνιση Δεδομένων συγκεντρωτικά	38
Εικόνα 4-10: Εμφάνιση operators	40
Εικόνα 4-11: Αποθήκευση δεδομένων μετά την πρώτη επεξεργασία	41
Εικόνα 4-12: Διαδικασία εύρεσης ακραίων τιμών	42
Εικόνα 4-13: Ορισμός παραμέτρων "Filter examples"	42
Εικόνα 4-14: Αποτελέσματα διαδικασίας εύρεσης ακραίων τιμών	43
Εικόνα 4-15: Εξαγωγή Δεδομένων σε μορφή .xlsx	44
Εικόνα 4-16: Αρχείο Matlab για εύρεση δείκτη Silhouette	56
Εικόνα 4-18: Σχεδιασμός της διαδικασίας εκτέλεσης για το Market Basket Analysis... ..	62

Κεφάλαιο 1 ΕΙΣΑΓΩΓΗ

1.1 ΣΚΟΠΟΣ ΤΗΣ ΕΡΓΑΣΙΑΣ

Αντικείμενο της παρούσας διπλωματικής εργασίας είναι η ανάλυση της συμπεριφοράς των καταναλωτών super markets με τη χρήση μεθόδων πολυκριτήριας ανάλυσης και μηχανικής μάθησης. Για την συλλογή δεδομένων συγκεντρώθηκαν ερωτηματολόγια μέσω διαδικτύου με χρήση ηλεκτρονικού ερωτηματολογίου, καθώς και με τη διαδικασία της δια ζώσης συνέντευξης σε ένα τυχαίο δείγμα καταναλωτών.

1.2. ΔΟΜΗ ΤΗΣ ΕΡΓΑΣΙΑΣ

Η εργασία απαρτίζεται από πέντε ενότητες. Στην πρώτη ενότητα γίνεται μία εισαγωγή της εργασίας. Στην δεύτερη ενότητα παρουσιάζονται έρευνες που αφορούν στην ανάλυση συμπεριφοράς καταναλωτών super markets (μέθοδοι ανάλυσης δεδομένων) και αναφέρεται ο στόχος και η διαδικασία που θα ακολουθηθεί στην παρούσα διπλωματική. Στην τρίτη ενότητα αναλύεται το θεωρητικό υπόβαθρο των μεθοδολογιών που χρησιμοποιήθηκαν στην συγκεκριμένη διπλωματική εργασία. Στην τέταρτη ενότητα αναλύεται ενδελεχώς η μεθοδολογία που χρησιμοποιήθηκε με πίνακες και διαγράμματα καθώς και τα αποτελέσματα αυτής. Στην πέμπτη ενότητα παρουσιάζονται τα συμπεράσματα της έρευνας.

Κεφάλαιο 2 ΥΦΙΣΤΑΜΕΝΗ ΚΑΤΑΣΤΑΣΗ

2.1. ΕΙΣΑΓΩΓΗ

Οι αγορές βασικών διατροφικών αγαθών, πλέον, πραγματοποιούνται στις μικρές και μεγάλες πόλεις κυρίως από super markets. Λόγω του σύγχρονου ρυθμού ζωής των κατοίκων των μεγάλων αστικών κέντρων, η αγορά όλων των προϊόντων που είναι απαραίτητα για καθημερινή χρήση (από τρόφιμα έως και είδη σπιτιού) από ένα μόνο κατάστημα είναι αναγκαστική επιλογή και ταυτόχρονα ανάγκη. Επειδή, λοιπόν, η αγορά από super market (υποκαταστήματα μεγάλων αλυσίδων ή και όχι) αφορά ένα μεγάλο κομμάτι, ίσως και το μεγαλύτερο ενός καταναλωτικού κοινού μίας πόλης, η έρευνα της συμπεριφοράς των καταναλωτών super market έχει εξετασθεί από διάφορες σκοπιές, τόσο των κριτηρίων με τα οποία αυτοί αγοράζουν, όσο και της συσχέτισης των προϊόντων σε κάθε συναλλαγή. Προηγούμενες έρευνες σχετικά με την ανάλυση της συμπεριφοράς των καταναλωτών υπεραγορών super markets διαφόρων χωρών παρατίθενται παρακάτω, καθώς και οι μέθοδοι που χρησιμοποιήθηκαν για την συλλογή και ανάλυση των δεδομένων.

2.2. ΥΦΙΣΤΑΜΕΝΗ ΚΑΤΑΣΤΑΣΗ

Σε έρευνα που διεξήχθη στην μητροπολιτική περιοχή της Κοπεγχάγης (Solgaard & Hansen, 2003) αναλύθηκαν οι προτιμήσεις των καταναλωτών για τρεις διαφορετικούς τύπους super markets που υπάρχουν στην Κοπεγχάγη και τα κριτήρια με τα οποία αυτοί επέλεξαν το εκάστοτε super market. Συλλέχθηκαν 631 ερωτηματολόγια και για την εξόρυξη των δεδομένων χρησιμοποιήθηκε το μοντέλο Logit, καθώς επίσης και το ιεραρχικό μοντέλο Bayes. Στο τέλος, τα αποτελέσματα και των δύο μοντέλων συγκρίθηκαν μεταξύ τους.

Σε άλλη έρευνα, (Aiguo, Ping, & Lanling, 2011) που δημοσιεύθηκε στο Διεθνές Συνέδριο CECNet χρησιμοποιήθηκαν διάφορες μέθοδοι για εξόρυξη δεδομένων. Συγκεκριμένα χρησιμοποιήθηκαν αλγόριθμοι συσχέτισης για να αναλυθεί η συσχέτιση μεταξύ των προϊόντων που αγοράζονται από καταναλωτές σε super markets (market basket analysis). Εξετάστηκε, επίσης, η σειρά με την οποία οι καταναλωτές αγοράζουν προϊόντα διαφορετικών ειδών και γι' αυτό χρησιμοποιήθηκε Sequence analysis. Τέλος, χρησιμοποιήθηκαν classification analysis για την περιγραφή και ταξινόμηση των πελατών, καθώς και αντίστροφα, cluster analysis προκειμένου να ταξινομηθούν οι πελάτες σε συγκεκριμένο αριθμό συστάδων.

Οι Ιακώβου, Καναβός και Τσακαλίδης (Iakovou, Kanavos, & Tsakalidis, 2016) στην έρευνά τους σχετικά με την μοντελοποίηση και πρόβλεψη συμπεριφοράς καταναλωτών ανέλυσαν τις αγορές κάθε καταναλωτή. Χρησιμοποιήθηκαν 10000 εγγραφές από συναλλαγές καταναλωτών σε περίοδο τεσσάρων χρόνων, οι οποίες

επεξεργάστηκαν μέσω 5 κατηγοριών ταξινομητών της βιβλιοθήκης της WEKA και συγκεκριμένα χρησιμοποιήθηκαν οι “Bayes”, “functions”, “lazy”, “rules” και “trees”.

Οι Akpınar, Ozkan, Ceylan και Sayin (Akpınar, Ozkan, Sayin, & Ceylan, 2010) στην έρευνά τους από 400 ερωτηματολόγια που συλλέχθηκαν τον Αύγουστο του 2004 εξέτασαν το προφίλ των καταναλωτών δύο super markets, με τη βοήθεια του SPSS, και εξέτασαν το ρίσκο του κάθε καταναλωτή.

Όσον αφορά στο *market basket analysis*, στην έρευνα των Yen-Liang Chen et al. (Chen, Tang, Shen, & Hu, 2004) εξετάστηκαν οι αγορές πελατών από διάφορα υποκαταστήματα ή καταστήματα αντιπροσώπων μίας μεγάλης εταιρίας, προκειμένου να εξεταστούν οι κανόνες συσχέτισης των προϊόντων της. Για την ανάλυση των δεδομένων χρησιμοποιήθηκε ένας προτεινόμενος, από τους συγγραφείς, αλγόριθμος συσχέτισης και συγκρίθηκαν τα αποτελέσματά του με τους κανόνες συσχέτισης του Apriori.

Τέλος, άλλη μία έρευνα των Troy Raeder και Nitesh V. Chawla (Raeder & Chawla, 2010) σχετικά με το *market basket analysis*, η οποία διενεργήθη από το Πανεπιστήμιο Notre Dame των ΗΠΑ, δημιουργεί ένα νέο τρόπο ανάλυσης για να εξετάσει τα προϊόντα που αγοράζονται συχνά μαζί. Δημιουργεί, δηλαδή, ένα δίκτυο προϊόντων, με τα προϊόντα να είναι οι κόμβοι και η συσχέτιση των προϊόντων που αγοράζονται μαζί να φαίνεται στις ενώσεις των κόμβων. Στην συνέχεια εφαρμόζονται κανόνες συσχέτισης στο γράφημα που έχει δημιουργηθεί βάζοντας όριο το πόσες φορές εμφανίζεται μία συγκεκριμένη σύνδεση μεταξύ δύο προϊόντων σε αυτό.

2.3. ΚΑΘΟΡΙΣΜΟΣ ΤΟΥ ΠΡΟΒΛΗΜΑΤΟΣ

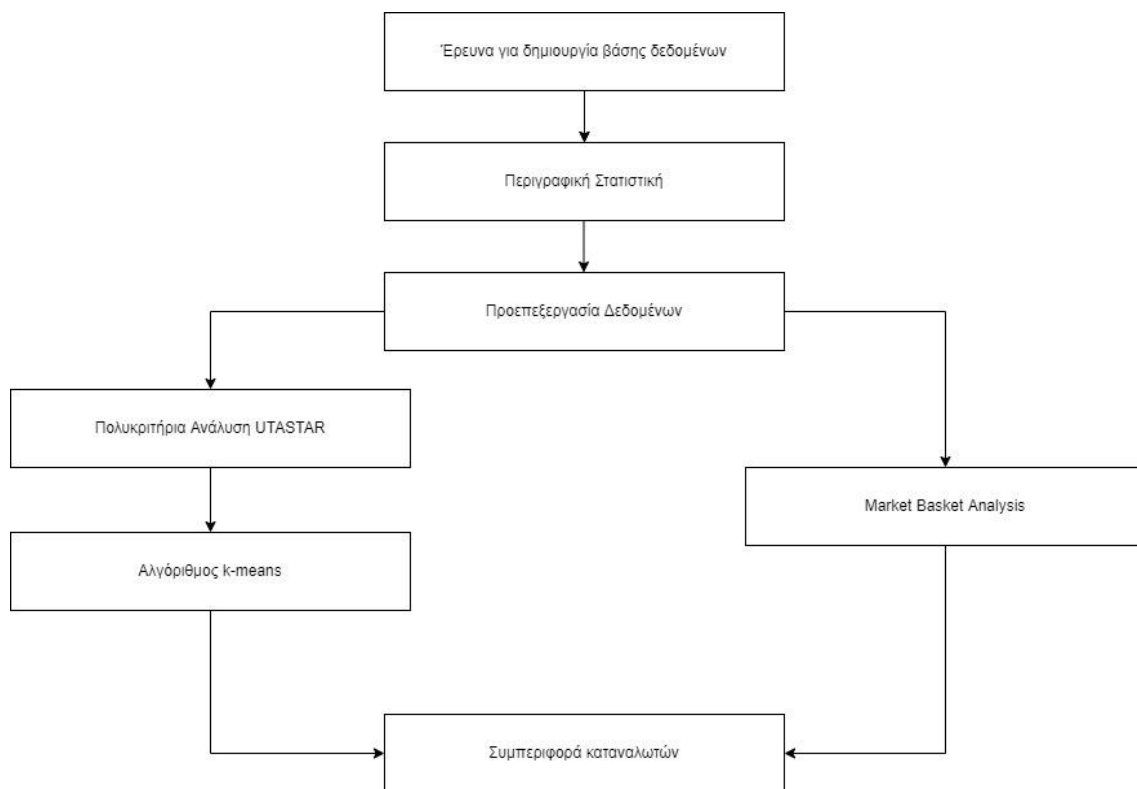
Στην παρούσα διπλωματική θα γίνει ανάλυση συμπεριφοράς καταναλωτή με χρήση μεθόδων πολυκριτήριας ανάλυσης και μηχανικής μάθησης. Τα δεδομένα αντλήθηκαν από ερωτηματολόγια που διανεμήθηκαν ηλεκτρονικά και δια ζώσης σε κατοίκους περιοχών του Δήμου Χανίων την περίοδο Ιανουαρίου-Μαρτίου 2022. Συνολικά συγκεντρώθηκαν 210 ερωτηματολόγια. Για την πολυκριτήρια ανάλυση χρησιμοποιήθηκε η μέθοδος *UTASTAR*. Αντίστοιχα, για την προεπεξεργασία των δεδομένων και την εφαρμογή των μεθόδων μηχανικής μάθησης, συγκεκριμένα του *market basket analysis* και των κανόνων συσχέτισης, χρησιμοποιήθηκε το λογισμικό *RapidMiner*. Για την συσταδοποίηση των κριτηρίων και των εναλλακτικών επιλογών των καταναλωτών έγινε χρήση του αλγορίθμου συσταδοποίησης *k-means* μέσω γλώσσας προγραμματισμού Matlab.

Κεφάλαιο 3 ΜΕΘΟΔΟΛΟΓΙΑ

3.1. ΕΙΣΑΓΩΓΗ

Σε αυτήν την ενότητα θα παρουσιαστεί μία ανάλυση του θεωρητικού υπόβαθρου των μεθόδων που ακολουθήθηκαν στην παρούσα διπλωματική εργασία. Στόχος της παρούσας διπλωματικής εργασίας είναι η ανάλυση συμπεριφοράς καταναλωτή για τα super markets του Ν. Χανίων.

Αρχικά, θα αναλυθεί η διαδικασία δημιουργίας μίας βάσης δεδομένων. Στη συνέχεια, τα δεδομένα θα χρησιμοποιηθούν ώστε να γίνει μία σύντομη στην περιγραφική στατιστική ανάλυση προκειμένου να δωθούν τα στατιστικά στοιχεία των ερωτώμενων. Τα ίδια δεδομένα θα υποστούν προεπεξεργασία και θα αναλυθεί με ποιες μεθόδους γίνεται η προεπεξεργασία δεδομένων και πιο συγκεκριμένα, ποια είναι τα προβλήματα που δημιουργούνται και με ποιο τρόπο αντιμετωπίζονται. Η προεπεξεργασία δεδομένων είναι απαραίτητη για την αποφυγή λανθασμένων και ασαφών συμπερασμάτων. Έπειτα, τα δεδομένα που προέρχονται από την προεπεξεργασία, θα χρησιμοποιηθούν ώστε να πραγματοποιηθεί η πολυκριτήρια ανάλυση UTASTAR, ώστε να εξαχθούν συμπεράσματα για τα βάρη των κριτηρίων με τα οποία οι καταναλωτές πραγματοποιούν τις αγορές τους, καθώς και την προτίμησή τους σε κάθε super market (ολικές χρησιμότητες). Τα αποτελέσματα της UTASTAR (τα οποία είναι τα βάρη των κριτηρίων και οι ολικές χρησιμότητες) θα χρησιμοποιηθούν ως είσοδοι για την δημιουργία συστάδων καταναλωτών μέσω του αλγόριθμου συσταδοποίησης K-means, της οποίας σκοπός είναι να δημιουργηθούν ομάδες καταναλωτών με κοινά χαρακτηριστικά, τα οποία θα εξεταστούν. Τέλος, τα δεδομένα από την προεπεξεργασία θα χρησιμοποιηθούν και για την εφαρμογή του Market Basket Analysis, προκειμένου να εξαχθούν οι κανόνες συσχέτισης των προϊόντων που οι καταναλωτές αγοράζουν μαζί από τα super markets. Από τις δύο αναλύσεις θα εξαχθούν τελικώς τα συμπεράσματα για την συμπεριφορά των καταναλωτών των super markets του Ν. Χανίων.



Σχήμα 3-1: Γράφημα ροής μεθοδολογίας

3.2. ΔΗΜΙΟΥΡΓΙΑ ΒΑΣΗΣ ΔΕΔΟΜΕΝΩΝ

Η καθολική πρόσβαση που παρέχει, στη σύγχρονη εποχή, το διαδίκτυο έχει θετικό αντίκτυπο και στην εύρεση δεδομένων. Η συλλογή δεδομένων και κατ' επέκταση, η δημιουργία μίας βάσης δεδομένων προς επεξεργασία μπορεί να πραγματοποιηθεί με ποικίλους τρόπους, όπως έτοιμες βάσεις δεδομένων από το διαδίκτυο, συλλογή ηλεκτρονικών ερωτηματολογίων, μέσω προσωπικών συνεντεύξεων και μέσω παρατήρησης. Στην συγκεκριμένη διπλωματική εργασία χρησιμοποιήθηκε η μέθοδος του ερωτηματολογίου, το οποίο διανεμήθηκε σε μέσα κοινωνικής δικτύωσης (CAWI- Web interviewing) και δια ζώσης, όπου αυτό ήταν εφικτό. Το ερωτηματολόγιο αποτελείται από μία σειρά δομημένων ερωτήσεων κλειστού και ανοιχτού τύπου, ίδιες για όλους τους ερωτώμενους.

3.3. ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ

Τα δεδομένα που αντλούνται από μία έρευνα πρέπει συχνά να υποβληθούν σε μία στατιστική ανάλυση προκειμένου να γίνουν κατανοητά από το ευρύ κοινό. Για αυτόν τον λόγο, χρησιμοποιούνται στατιστικές παράμετροι, προκειμένου τα δεδομένα να αναλυθούν επαρκώς και κατ' επέκταση να οπτικοποιηθούν για την καλύτερη κατανόησή τους. Η περιγραφική στατιστική έχει ως στόχο την συνοπτική οργάνωση και παρουσίαση των δεδομένων μίας έρευνας. Στην περιγραφική στατιστική

χρησιμοποιούνται συχνά δείκτες, όπως η διάμεσος και η αποκλίση. (Landau & Everitt , 2004)

3.4. ΠΡΟΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ

Η προεπεξεργασία των δεδομένων είναι από τις σημαντικότερες διαδικασίες που θα πρέπει να κάνει ένας ερευνητής προτού προχωρήσει στην οποιαδήποτε ανάλυση δεδομένων, καθώς αποτρέπει από εξαγωγή λανθασμένων συμπερασμάτων. Τα δεδομένα που λαμβάνονται αυτούσια από μία έρευνα δύναται να έχουν πολλών ειδών προβλήματα. Τα δεδομένα, πριν την προεπεξεργασία, καλούνται «ακάθαρτα» και η διαδικασία αντιμετώπισης των προβλημάτων τους ονομάζεται «καθαρισμός δεδομένων». Η προεπεξεργασία των δεδομένων περιλαμβάνει τον καθαρισμό τους, αλλά και τον μετασχηματισμό τους. Η διακριτοποίηση και η κανονικοποίηση είναι από τις πιο συνηθισμένες διαδικασίες της προεπεξεργασίας δεδομένων. Η διακριτοποίηση πραγματοποιείται με τον μετασχηματισμό αριθμητικών τιμών σε ονομαστικές τιμές, ενώ η κανονικοποίηση είναι η μετατροπή της κατανομής των αριθμητικών τιμών σε εύρος τιμών μεταξύ 0 και 1. Ακόμα ένα θέμα που συγκαταλέγεται στην προεπεξεργασία των δεδομένων είναι η μείωση του όγκου τους. Αναλυτικότερα τα προβλήματα δεδομένων και οι τρόποι αντιμετώπισης αναφέρονται παρακάτω.

3.4.1. Καθαρισμός Δεδομένων

Ο καθαρισμός δεδομένων πραγματοποιείται με τις παρακάτω ενέργειες:

- συμπλήρωση ελλিপών τιμών,
- εύρεση ακραίων τιμών (outliers) και εξομάλυνση, εφόσον περιέχουν θόρυβο,
- διόρθωση ασυνεπειών.

3.4.1.1. Ελλιπείς τιμές

Συχνά παρατηρείται έλλειψη κάποιων τιμών στα δεδομένα, που σημαίνει ότι τα δεδομένα δεν είναι κατάλληλα προς επεξεργασία διότι οι ελλείψεις αυτές επηρεάζουν τις τιμές των αποτελεσμάτων. Αυτές είναι οι ελλιπείς τιμές. Το φαινόμενο αυτό μπορεί να οφείλεται σε κάποια δυσλειτουργία του συστήματος αποθήκευσης δεδομένων (αν αποθηκεύονται ψηφιακά), σε ασυμβατότητα με άλλα δεδομένα, που οδήγησε στη διαγραφή ή στη μη καταχώρηση τους. Επομένως, για την συμπλήρωση των ελλিপών δεδομένων αρχικά, πρέπει να αναγνωριστούν οι γραμμές που τα περιέχουν. Αν οι ελλιπείς τιμές αποτελούν ένα μικρό πληθος συγκριτικά με το σύνολο των δεδομένων, τότε αυτές διαγράφονται. Διαφορετικά, ακολουθείται κάποια από τις παρακάτω διαδικασίες προκειμένου να συμπληρωθούν.

Μερικές από τις αυτοματοποιημένες μεθόδους συμπλήρωσης ελλιπών τιμών είναι οι παρακάτω:

- Χρησιμοποιείται μία σταθερά π.χ. «-1», «unknown». Συμπληρώνεται, δηλαδή, με μία τιμή η οποία δεν επηρεάζει το σύνολο ή συναντάται συχνά.
- Αντικαθίστανται από την τιμή της μέσης τιμής του συγκεκριμένου χαρακτηριστικού, αν είναι αριθμητικό, διαφορετικά με την πιο συνηθισμένη τιμή, αν το χαρακτηριστικό είναι ονομαστικό.
- Αντικατάσταση με την πλέον εμφανιζόμενη τιμή, αν προκειται για κατηγορική μεταβλητή.

3.4.1.2. Δεδομένα με θόρυβο

Δεδομένα με θόρυβο εννοούνται τα δεδομένα τα οποία περιέχουν ακραίες τιμές ή λανθασμένες τιμές λόγω κάποιου προβλήματος στην καταγραφή τους από ανθρώπινη αμέλεια ή λάθος μέτρησης. Υπάρχουν δύο τεχνικές με τις οποίες αντιμετωπίζονται τα δεδομένα με θόρυβο. Στην πρώτη αντικαθίστουμε όλες οι τιμές, με άλλες καταλληλότερες. Στην δεύτερη, αρχικά οι ακραίες τιμές εντοπίζονται και στη συνέχεια, είτε διαγράφονται, είτε μετατρέπονται.

Μία μέθοδος η οποία χρησιμοποιείται είναι αυτή της ενδοχείωσης. Τα δεδομένα τοποθετούνται σε άυξουσα σειρά και χωρίζονται σε δοχεία. Ο χωρισμός σε δοχεία γίνεται είτε με βάση το πλάτος είτε με βάση τη συχνότητα. Τα δοχεία ίσου πλάτους έχουν όλα το ίδιο διάστημα τιμών, ενώ τα δοχεία ίδιας συχνότητας έχουν όλα ίδιο πλήθος. Αφού βρεθούν τα δοχεία, τότε υπολογίζονται νέες τιμές για κάθε δοχείο και αντικαθιστούν αυτές που έχουν θόρυβο. Οι νέες τιμές δύναται να είναι είτε η μέση τιμή του κάθε δοχείου, είτε μία οριακή τιμή. Σχετικά με την αντικατάσταση με οριακή τιμή, αυτό εξαρτάται από την ακραία τιμή που αντικαθίσταται. Αν βρίσκεται κοντά σε κάποια από τις δύο οριακές τιμές, τότε αντικαθίσταται με αυτήν.

Μία δεύτερη μέθοδος, αρκετά διαδεδομένη είναι αυτή της συσταδοποίησης στην οποία τα δεδομένα συγκεντρώνονται σε συστάδες με βάση την ομοιότητά τους. Τα δεδομένα που δεν ανήκουν σε καμία συστάδα αποτελούν τα δεδομένα με θόρυβο.

3.4.1.3. Ασυνεπή Δεδομένα

Η ύπαρξη ασυνέπειας στα δεδομένα προκαλείται από ανθρώπινο σφάλμα κατά την εισαγωγή των δεδομένων, εσκεμμένο σφάλμα (π.χ. όταν ο χρήστης δεν θέλει να αποκαλύψει προσωπικές πληροφορίες) καθώς και από κακό σχεδιασμό της καταχώρησης δεδομένων (π.χ. πολλά προαιρετικά πεδία). Πιθανές αιτίες είναι ο διαφορετικός τρόπος αναπαράστασης ή και η χρήση διαφορετικών κλιμάκων, π.χ. μονάδες μέτρησης, διαφορετικό νόμισμα. Για την διόρθωση του προβλήματος

συνήθως χρησιμοποιούνται κατάλληλα λογισμικά καθαρισμού δεδομένων (π.χ. ορθογραφικού ελέγχου).

3.4.2. Μετασχηματισμός και Διακριτοποίηση Δεδομένων

Με τον μετασχηματισμό των δεδομένων επιτυγχάνεται η σύγκριση δεδομένων που αρχικά είναι ανόμοια και είναι αδύνατον να συγκριθούν. Επίσης, ταυτόχρονα μπορεί να επιτευχθεί και η μείωση του όγκου των δεδομένων με την μετατροπή των τιμών ενός χαρακτηριστικού σε κάποιο υποσύνολό τους ή με την κανονικοποίηση των τιμών ενός χαρακτηριστικού. Η διακριτοποίηση μπορεί να θεωρηθεί υποκατηγορία του μετασχηματισμού δεδομένων, η οποία συμβαίνει με την μετατροπή ενός διαστήματος τιμών σε διακριτές τιμές.

3.4.3. Μείωση Δεδομένων

Με την μείωση δεδομένων αντιμετωπίζεται το συχνό πρόβλημα του μεγάλου όγκου δεδομένων προς ανάλυση, το οποίο συνήθως απαιτεί χρόνο και ικανό εξοπλισμό. Η μείωση δεδομένων στοχεύει σε μία μείωση του όγκου που να παράγει, όμως ίδια ή αντίστοιχα αποτελέσματα. Δυο τεχνικές, με τις οποίες μπορεί να επιτευχθεί μείωση των διαστάσεων, είναι η επιλογή χαρακτηριστικών και ο μετασχηματισμός των δεδομένων. Με την τεχνική της επιλογής χαρακτηριστικών επιλέγεται το ελάχιστο πλήθος χαρακτηριστικών, με τα οποία είναι εφικτό να παραχθούν ισοδύναμα ή όσο το δυνατόν κοντινότερα αποτελέσματα με αυτά που θα παίρναμε, αν χρησιμοποιούσαμε όλα τα χαρακτηριστικά για ανάλυση.

3.5. ΠΟΛΥΚΡΙΤΗΡΙΑ ΑΝΑΛΥΣΗ UTASTAR

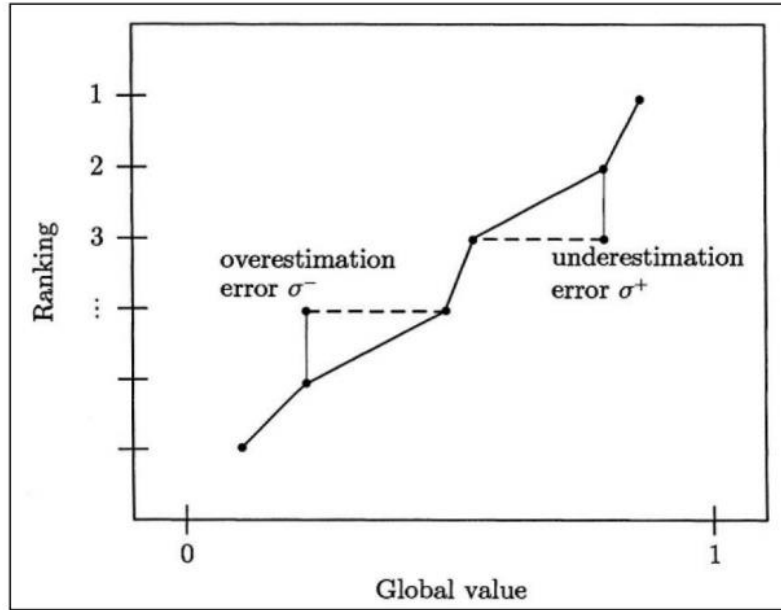
3.5.1. Εισαγωγή

Η ανάγκη των επιχειρήσεων να κατανοήσουν και να αναλύσουν τον τρόπο με τον οποίο οι καταναλωτές αποφασίζουν, οδήγησε αρχικά στην δημιουργία ενός μοντέλου πολυκριτήριας ανάλυσης που στόχο είχε την αναπαράσταση των βαρών των κριτηρίων απόφασης κάθε καταναλωτή.

3.5.2. Μέθοδος UTASTAR

Η μέθοδος UTASTAR προτάθηκε από τους Siskos-Yannacopoulos (1985) και αποτελεί μια βελτιωμένη έκδοση της μεθόδου UTA των Jacquet-Lagrece και Siskos, 1982 (Siskos & Yannacopoulos, 1985) (Siskos, Grigoroudis, & Matsatsinis, 2016), στην οποία αρχική έκδοση η βασική ιδέα είναι πως για κάθε μία δράση $a \in A$ ορίζεται ένα μοναδικό σφάλμα $\sigma(a)$. Αυτή η συνάρτηση σφάλματος δεν είναι επαρκής για την ελαχιστοποίηση της

ολικής διασποράς των σημείων στη μονότονη καμπύλη του σχήματος 3-2. Το πρόβλημα αφορά τα σημεία που βρίσκονται δεξιά της καμπύλης, από τα οποία θα ήταν προτιμότερο να αφαιρεθεί μια ποσότητα αξίας χωρίς να αυξηθούν οι αξίες των άλλων.



Σχήμα 3-2: Καμπύλη μονότονης παλινδρόμησης (Siskos & Yannacopoulos, 1985)

Σε αντίθεση με την UTA, στην οποία υπάρχει μονή συνάρτηση σφάλματος, στην UTASTAR εισάγεται μία διπλή συνάρτηση σφάλματος και ο τύπος γίνεται:

$$U'(g(a)) = \sum_{i=1}^n u_i(g_i(a)) - \sigma^+(a) + \sigma^-(a) \quad \forall a \in A \quad (3.1)$$

Όπου, $\sigma^+(a)$ και $\sigma^-(a)$ τα σφάλματα υποεκτίμησης και υπερκεκτίμησης, αντίστοιχα.

Ακόμα μία σημαντική τροποποίηση αφορά στους περιορισμούς της μονοτονίας των κριτηρίων, ο οποίοι μοντελοποιούνται με τον ακόλουθο τρόπο μετασχηματισμού των μεταβλητών:

$$W_{i,j} = u_i(g_i^j) \geq 0 \quad \forall i = 1, 2, \dots, n \text{ και } j = 1, 2, \dots, a_i - 1 \quad (3.2)$$

Με αυτόν τον τρόπο οι περιορισμοί μη αρνητικότητας αντικαθιστούν της συνθήκες μονοτονίας.

Συνοπτικά ο αλγόριθμος της UTASTAR έχει ως εξής:

Βήμα 1: Η ολική αξία των εναλλακτικών $U(g(a_k))$, $k = 1, 2, \dots, m$ εκφράζεται αρχικά ως συνάρτηση των περιθωρίων αξιών $u_i(g_i)$ και στη συνέχεια των μεταβλητών $W_{i,j}$ από την εξίσωση (3.2), μέσω των ακόλουθων σχέσεων:

$$u_i(g_i^1) = 0, \forall i = 1, 2, \dots, n$$

$$u_i(g_i^1) = \sum_{t=1}^{j-1} w_{it}, \forall i = 1, 2, \dots, n \text{ και } j = 2, 3, \dots, a_i - 1 \quad (3.3)$$

Βήμα 2: Γίνεται εισαγωγή δύο συναρτήσεων σφάλματος $\sigma^+(a)$ και $\sigma^-(a)$ στο A, και για κάθε ζεύγари διαδοχικών δράσεων στην προδιάταξη γράφονται οι αναλυτικές εκφράσεις:

$$\Delta(a_k, a_{k+1}) = U(g(a_k)) - \sigma^+(a_k) + \sigma^-(a_k) - U(g(a_{k+1})) + \sigma^+(a_{k+1}) - \sigma^-(a_{k+1}) \quad (3.4)$$

Βήμα 3: Δημιουργείται το γραμμικό πρόβλημα προς επίλυση, όπως παρουσιάζεται παρακάτω:

$$[min]z = \sum_{k=1}^m \sigma^+(a_k) + \sigma^-(a_k)$$

Υπό τους περιορισμούς

$$\Delta(a_k, a_{k+1}) \geq \delta \text{ εάν } a_k > a_{k+1}, \forall k$$

$$\Delta(a_k, a_{k+1}) = 0 \text{ εάν } a_k \sim a_{k+1}, \forall k$$

$$\sum_{i=1}^n \sum_{j=1}^{a_i-1} w_{ij} = 1$$

$$w_{ij} \geq 0, \quad \sigma^+(a_k) \geq 0, \quad \sigma^-(a_k) \geq 0, \forall i, j \text{ και } k \quad (3.5)$$

Όπου δ ένας μικρός θετικός αριθμός.

Βήμα 4: Γίνεται έλεγχος για την ύπαρξη πολλαπλών βέλτιστων ή ημιβέλτιστων λύσεων στο γραμμικό πρόβλημα (3.5), με τον υπολογισμό του βαρύκεντρου των προσθετικών συναρτήσεων αξίας που μεγιστοποιούν τις παρακάτω αντικειμενικές συναρτήσεις:

$$u_i(g_i^*) = \sum_{j=1}^{a_i-1} w_{ij}, \forall i = 1, 2, \dots, n \quad (3.6)$$

Στο σύνολο των περιορισμών του παραπάνω γραμμικού προβλήματος (3.5) προστίθεται και ο παρακάτω περιορισμός:

$$\sum_{k=1}^m (\sigma^+(a_k) + \sigma^-(a_k)) \leq z^* + \varepsilon \quad (3.7)$$

Όπου z^* είναι η βέλτιστη τιμή (σφάλμα) του γραμμικού προβλήματος του βήματος 3 και ε ένας πολύ μικρός θετικός αριθμός ή μηδέν.

Αποδείχθηκε από τους Siskos-Yannacopoulos σε ένα σύνολο πειραματικών δεδομένων, ότι η UTASTAR δίνει καλύτερα αποτελέσματα από την UTA.

3.6. ΣΥΣΤΑΔΟΠΟΙΗΣΗ

3.6.1. Εισαγωγή

Συσταδοποίηση είναι η διαδικασία κατά την οποία τα δεδομένα χωρίζονται σε ένα σύνολο ομάδων. Στόχος της δημιουργίας συστάδων (Clusters) είναι η ομαδοποίηση όσο το δυνατόν πιο όμοιων δεδομένων μαζί και αντίστοιχα η έντονη διαφοροποίηση των χαρακτηριστικών της μίας συστάδας από την άλλη. Η κατηγοριοποίηση των μεθόδων συσταδοποίησης γίνεται με βάση:

- τον τύπο των δεδομένων που προορίζονται για συσταδοποίηση,
- την μέθοδο που χρησιμοποιείται για την συσταδοποίηση των δεδομένων και
- τη θεωρία και τις θεμελιώδεις έννοιες στις οποίες είναι βασισμένες οι τεχνικές ανάλυσης της συστάδας.

Οι πιο γνωστοί αλγόριθμοι συσταδοποίησης είναι: K-means, OPTICS, DBSCAN κ.α. (Ματσατσίνης, 2021)

3.6.2. Συσταδοποίηση μέσω MARKEX

Το λογισμικό MARKEX δίνει την δυνατότητα ομαδοποίησης των ερωτώμενων. Η συσταδοποίηση πραγματοποιείται με την χρήση των βαρών των κριτηρίων, τα οποία προκύπτουν από την δημιουργία των προσθετικών χρησιμότητων μέσω της UTASTAR. Στον χρήστη δίνεται η δυνατότητα ανάλογα με τις επιλογές που θα κάνει στην σημαντικότητα των κριτηρίων να δημιουργήσει συστάδες καταναλωτών που εναρμονίζονται με την αντίστοιχη επιλογή σημαντικότητας κριτηρίων. Συγκεκριμένα, όσον αφορά τη σημαντικότητα των κριτηρίων πρέπει να σημειωθούν τα εξής:

- Για κριτήριο που θα θεωρηθεί σημαντικό, τότε το κατώφλι του βάρους που επιλέγεται, θεωρείται ελάχιστο.
- Για κριτήριο που δεν θεωρείται σημαντικό, το κατώφλι του βάρους που επιλέγεται, θεωρείται μέγιστο.
- Τέλος, για κριτήριο το οποίο θεωρείται αδιάφορο, δεν ορίζεται κάποιο κατώφλι.

3.6.3. Αλγόριθμος K-means

Στόχος του αλγορίθμου k-means είναι η ελαχιστοποίηση της μέσης τετραγωνικής απόστασης των δειγμάτων από τα κέντρο της συστάδας στην οποία ανατίθενται.

Ορίζεται από τον χρήστη ο αριθμός των συστάδων k . Ο αλγόριθμος k-means ξεκινάει με k τυχαία σημεία, τα οποία ονομάζονται κεντροειδή της συστάδας και δηλώνουν το κέντρο βάρους της συστάδας. Αντιστοιχίζονται με τον αντιπροσωπευτικότερο καταναλωτή του τμήματος αγοράς.

Ο αλγόριθμος εκτελείται σε 4 επαναληπτικά βήματα:

1. Διαχωρισμός των δεδομένων σε τυχαία και μη κενά σύνολα.
2. Υπολογίζονται τα σημεία που θα είναι τα κέντρα βάρους των συστάδων του τρέχοντος διαχωρισμού.
3. Γίνεται εναπόθεση του κάθε σημείου στη συστάδα με το κοντινότερο κεντρικό σημείο.
4. Παίρνοντας τον μέσο όρο των δειγμάτων κάθε συστάδας, επαναυπολογίζονται τα κέντρα των συστάδων, έτσι ώστε το κάθε κέντρο συστάδας να είναι το κοντινότερο στα σημεία που ανήκουν στη συστάδα του.

Τα βήματα αυτά εκτελούνται επαναληπτικά μέχρις ότου τα κέντρα των συστάδων να αλλάζουν από ελάχιστα έως και σε απόσταση μικρότερη από κάποια δοθείσα τιμή κατωφλίου. Εναλλακτικά, ο αλγόριθμος δύναται να τερματίσει όταν φτάσει ένα συγκεκριμένο αριθμό επαναλήψεων που θα δοθεί από τον χρήστη.

Συχνά, η τυχαία αρχικοποίηση των κέντρων των συστάδων επηρεάζει κατα πολύ και αρνητικά το αποτέλεσμα της συσταδοποίησης. Γι' αυτό τον λόγο, χρησιμοποιείται ένας δείκτης, ο Silhouette, ο οποίος μετράει την ποιότητα της συσταδοποίησης. Αυτός ο δείκτης μετράει την ομοιομορφία της κάθε συστάδας και την ανομοιομορφία των συστάδων μεταξύ τους, δηλαδή χρησιμοποιείται για τον χαρακτηρισμό της ποιότητας μιας ομαδοποίησης χωρίς επίβλεψη. Κυμαίνεται από -1 έως $+1$, όπου μία τιμή κοντά στο $+1$ δείχνει ότι το δείγμα ταιριάζει απόλυτα με την συστάδα στην οποία ανήκει, ενώ διαφέρει πολύ από τις γειτονικές συστάδες. Αν πάρει την τιμή -1 , τότε το δείγμα πρέπει να τοποθετηθεί στην γειτονική συστάδα. Σε περίπτωση που ο δείκτης πάρει την τιμή 0 , τότε το δείγμα που αναλύεται βρίσκεται σε ίδια απόσταση μεταξύ δύο γειτονικών συστάδων.

Ο συντελεστής Silhouette για κάθε σημείο i μίας συστάδας, υπολογίζεται ως εξής:

Υπολογισμός a_i : η μέση απόσταση του i από τα σημεία της συστάδας του.

Υπολογισμός b_i : η μέση απόσταση του i από όλα τα σημεία κάθε άλλης συστάδας, επιλογή του μικρότερου, δηλαδή μέση απόσταση από την κοντινότερη ομάδα.

Έτσι ο δείκτης Silhouette υπολογίζεται, όπως παρακάτω:

$$S_i = \begin{cases} 1 - \frac{a_i}{b_i}, & \text{εάν } a_i < b_i \\ 0, & \text{εάν } a_i = b_i \\ \frac{b_i}{a_i} - 1, & \text{εάν } a_i > b_i \end{cases} \quad (3.8)$$

Ο μέσος S_i του συνόλου δεδομένων δείχνει κατά πόσο ορθά ομαδοποιημένο είναι ένα σημείο (James et al, 2013).

3.7. MARKET BASKET ANALYSIS

3.7.1. Εισαγωγή

Το Market Basket Analysis (MBA) είναι μία από τις γνωστότερες μεθόδους μηχανικής μάθησης, η οποία χρησιμοποιείται ευρέως στον τομέα του εμπορίου. Αναλύοντας τις αγορές και τις προτιμήσεις των καταναλωτών, δύναται να ανιχνευθούν τα μοτίβα που χρησιμοποιούν, υποσυνείδητα ή μη, οι καταναλωτές για να πραγματοποιήσουν τις αγορές τους. Χρησιμοποιεί δηλαδή, τις συσχετίσεις που υπάρχουν στα προϊόντα που αγοράζουν συχνά οι καταναλωτές μαζί. Για την εφαρμογή του MBA χρησιμοποιούνται κυρίως αλγόριθμοι συσχέτισης, οι οποίοι ανακαλύπτουν τους ισχυρούς κανόνες συσχέτισης που «κρύβονται» στις συναλλαγές των καταναλωτών.

3.7.2. Market Basket Analysis

Η εξόρυξη συνόλων δεδομένων οδηγεί συχνά στην ανακάλυψη συσχετίσεων μεταξύ των δεδομένων. Με την πληθώρα δεδομένων που υπάρχουν, πολλές επιχειρήσεις δείχνουν ενδιαφέρον στην εξαγωγή μοτίβων από τα δεδομένα. Η εξαγωγή κανόνων συσχέτισης βοηθάει στις διαδικασίες λήψης αποφάσεων, όπως η ανάλυση συμπεριφοράς καταναλωτών. Το MBA προσδιορίζει τις αγοραστικές συνήθειες των καταναλωτών, βρίσκοντας συσχετίσεις μεταξύ των διαφορετικών ειδών που τοποθετούν οι πελάτες στα «καλάθια αγορών» τους. Η εξόρυξη αυτών των συσχετίσεων βοηθάει στην υλοποίηση στρατηγικών μαρκετινγκ τις επιχειρήσεις.

Αλγόριθμοι που χρησιμοποιούνται συχνά στο MBA:

- AIS (Artificial Immune System)
- SETM Algorithm
- Apriori Algorithm
- FP Growth

Στο MBA, κάθε αντικείμενο αντικαθίσταται από μία μεταβλητή που αντιπροσωπεύει την ύπαρξη ή όχι του συγκεκριμένου αντικειμένου μέσα σε ένα καλάθι αγορών. Επίσης, κάθε καλάθι αγορών αναπαρίσταται από ένα διάνυσμα που περιέχει τις μεταβλητές των αντικειμένων των αγορών. Τα διανύσματα αυτά αναλύονται από μοτίβα που αντικατοπτρίζουν τα προϊόντα που έχουν αγοραστεί μαζί ή που σχετίζονται συχνά μεταξύ τους.

Θεωρείται αρχικά ένα σύνολο προϊόντων $I = \{I_1, I_2, \dots, I_m\}$ και D θεωρείται μία βάση δεδομένων που περιέχονται οι συναλλαγές, καθε μία από τις οποίες συμβολίζεται με T και για τις οποίες ισχύει: $T \subseteq I$

Κάθε συναλλαγή σχετίζεται με ένα αναγνωριστικό, το οποίο ονομάζεται TID. Έστω A ένα σύνολο δεδομένων, το οποίο περιέχεται σε μία συναλλαγή, δηλαδή ισχύει η σχέση: $A \subseteq T$.

Ένας κανόνας συσχέτισης είναι μία σχέση της μορφής:

$$A \Rightarrow B, \text{ όπου } A \subset I, B \subset I \text{ και } A \cap B = \varnothing$$

Ο κανόνας $A \Rightarrow B$ ισχύει στο σύνολο των δεδομένων D με υποστήριξη (support) “ s ”, όπου “ s ” είναι το ποσοστό των συναλλαγών που περιέχουν $A \cup B$, δηλαδή τα στοιχεία του συνόλου A και του συνόλου B ή και των δύο μαζί. Αυτό λαμβάνεται ως πιθανότητα $P(A \cup B)$.

Ο κανόνας $A \Rightarrow B$ έχει εμπιστοσύνη (confidence) “ c ”, όπου “ c ” είναι το ποσοστό των συναλλαγών στο D που περιέχει το B υπό την προϋπόθεση να περιέχει και το A . Αυτό λαμβάνεται ως πιθανότητα $P(B|A)$.

Αυτό μεταφράζεται ως:

$$Support(A \Rightarrow B) = P(A \cup B)$$

$$Confidence(A \Rightarrow B) = P(B|A)$$

Οι κανόνες που ικανοποιούν ταυτόχρονα τόσο ένα ελάχιστο όριο υποστήριξης, όσο και ένα ελάχιστο όριο εμπιστοσύνης, ονομάζονται «Ισχυροί».

Συχνά χρησιμοποιείται και το κριτήριο “Lift” για την ανίχνευση της τυχαιότητας των κανόνων συσχέτισης. Το lift είναι ένα μέτρο το οποίο δείχνει κατα πόσο χειρότερη ή καλύτερη είναι η επίδοση ενός κανόνα σε σχέση με έναν κανόνα τυχαίας επιλογής.

$$Lift(X \rightarrow Y) = \frac{supp(X \rightarrow Y)}{supp(X) * supp(Y)}$$

Αν η τιμή του lift είναι κοντά στο 1, τότε δεν υπάρχει συσχέτιση μεταξύ του X και του Y και ο κανόνας εκφράζει ένα τυχαίο γεγονός. Αν η τιμή του lift είναι μεγαλύτερη από 1,

τότε υπάρχει θετική συσχέτιση μεταξύ των X και Y και η εμφάνιση του X ενισχύει την πιθανότητα εμφάνισης του Y . Αν το lift είναι μικρότερο του 1, τότε η συσχέτιση είναι αρνητική και η εμφάνιση του X μειώνει την πιθανότητα εμφάνισης του Y .

Εν κατακλείδι, η εξόρυξη δεδομένων μέσω κανόνων συσχέτισης μπορεί να συνοψισθεί σε δύο βήματα.

Βήμα 1^ο : Εύρεση όλων των συχνών συνόλων δεδομένων. Εξ' ορισμού, κάθενα από αυτά τα σύνολα στοιχείων θα εμφανίζεται τουλάχιστον τόσο συχνά, όσο ο προκαθορισμένος αριθμός υποστήριξης.

Βήμα 2^ο : Δημιουργία κανόνων συσχέτισης από τα σύνολα των συχνών στοιχείων. Αυτοί οι κανόνες πρέπει να ικανοποιούν το ελάχιστο όριο υποστήριξης και εμπιστοσύνης.

(Ματσατσίνης, Επιχειρηματική Ευφυΐα, Επιχειρηματική Αναλυτική και Ανάλυση Μεγάλων Δεδομένων, 2021)

Κεφάλαιο 4 ΑΠΟΤΕΛΕΣΜΑΤΑ

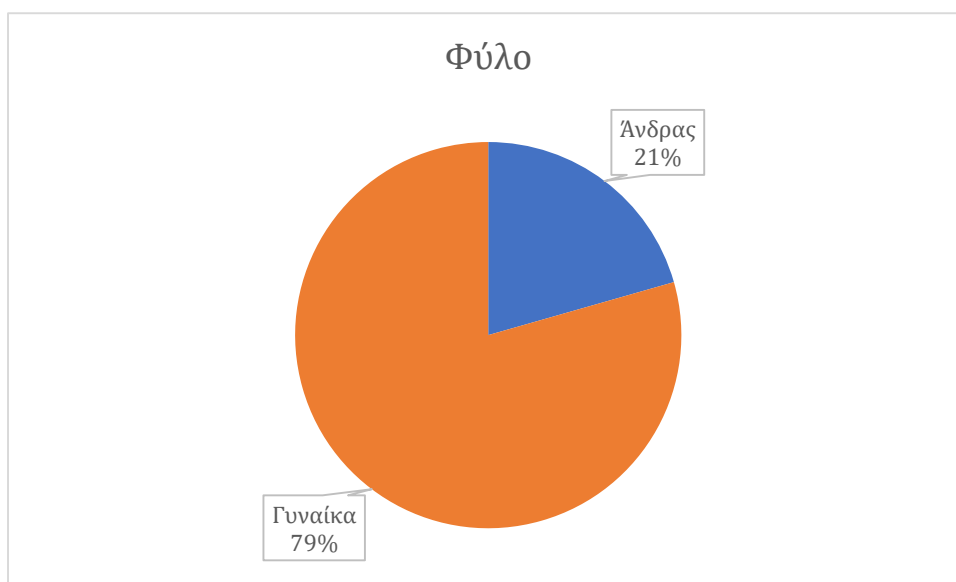
4.1. ΕΙΣΑΓΩΓΗ

Για την εκπόνηση της παρούσας διπλωματικής συλλέχθηκαν δεδομένα με την μέθοδο του ερωτηματολογίου. Τα ερωτηματολόγια διανεμήθηκαν τόσο ηλεκτρονικά, στα μέσα κοινωνικής δικτύωσης, όσο και σε έντυπη μορφή. Η έρευνα διεξήχθη την περίοδο Ιανουαρίου-Μαρτίου 2022 και συνολικά συλλέχθηκαν 210 ερωτηματολόγια από κατοίκους του Δήμου Χανίων. Το ερωτηματολόγιο αποτελούνταν από δημογραφικές ερωτήσεις, ερωτήσεις κατανόησης των προτιμήσεων του καταναλωτή, ενώ στη συνέχεια παρατίθονταν ερωτήσεις πολυκριτήριων επιλογών, καθώς και μία κατάταξη των εναλλακτικών super markets από τον καταναλωτή.

4.2. ΠΕΡΙΓΡΑΦΙΚΗ ΣΤΑΤΙΣΤΙΚΗ

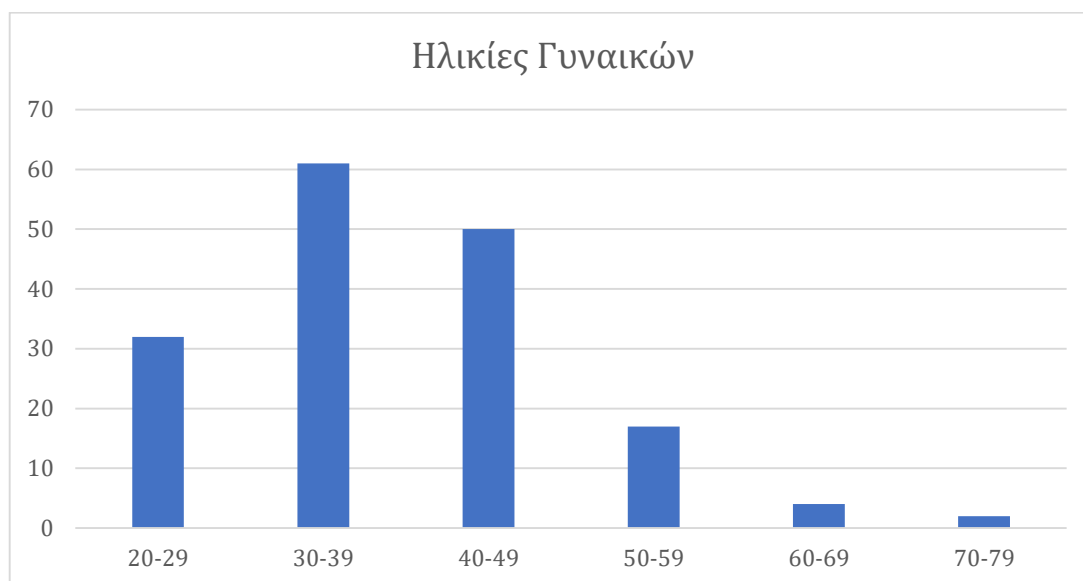
Σε αυτό το υποκεφάλαιο θα γίνει στατιστική ανάλυση των δεδομένων του δείγματος.

Οι παρατηρήσεις είναι σε σύνολο 210. Από αυτές, 166 παρατηρήσεις αντιστοιχούν σε γυναίκες και 43 σε άνδρες. Αυτό αντιστοιχεί σε ποσοστό 79% για τις γυναίκες και 21% για τους άνδρες. Παρατηρείται ότι 1 παρατήρηση δεν έχει τιμή και γι' αυτό το άθροισμα ανδρών και γυναικών δεν ισούται με 210.



Γράφημα 4-1: Ποσοστιαία κατανομή φύλων

Στη συνέχεια παρουσιάζεται η ηλικιακή κατανομή ανδρών και γυναικών σε δύο διαφορετικά γραφήματα.



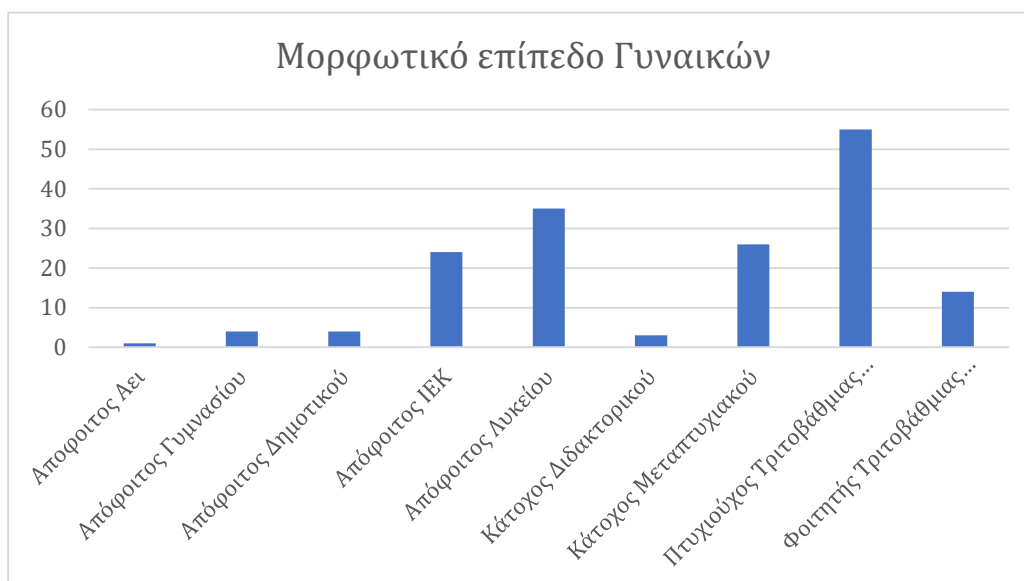
Γράφημα 4-2: Ηλικιακή κατανομή γυναικών

Οι επικρατέστερες ηλικιακές ομάδες στις γυναίκες είναι η 30-39 και ακολουθεί αυτή των 40-49.

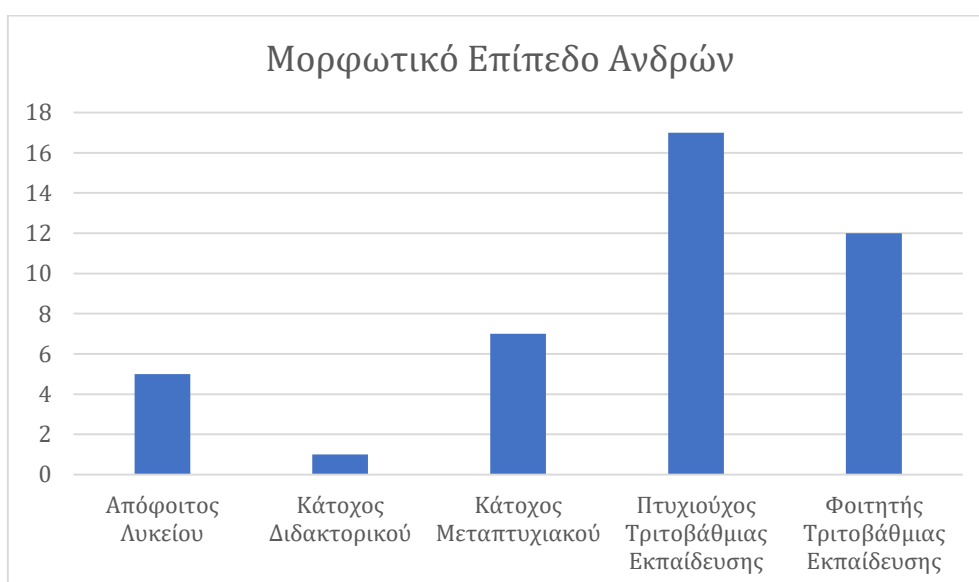


Γράφημα 4-3: Ηλικιακή κατανομή ανδρών

Στις ηλικίες των ανδρών η μεγαλύτερη είναι η ηλικιακή ομάδα των 20-29 και ακολουθεί αυτή των 30-39, στη συνέχεια αυτή των 50-60 και τέλος η ηλικιακή ομάδα των 40-49.



Γράφημα 4-4: Μορφωτικό επίπεδο γυναικών

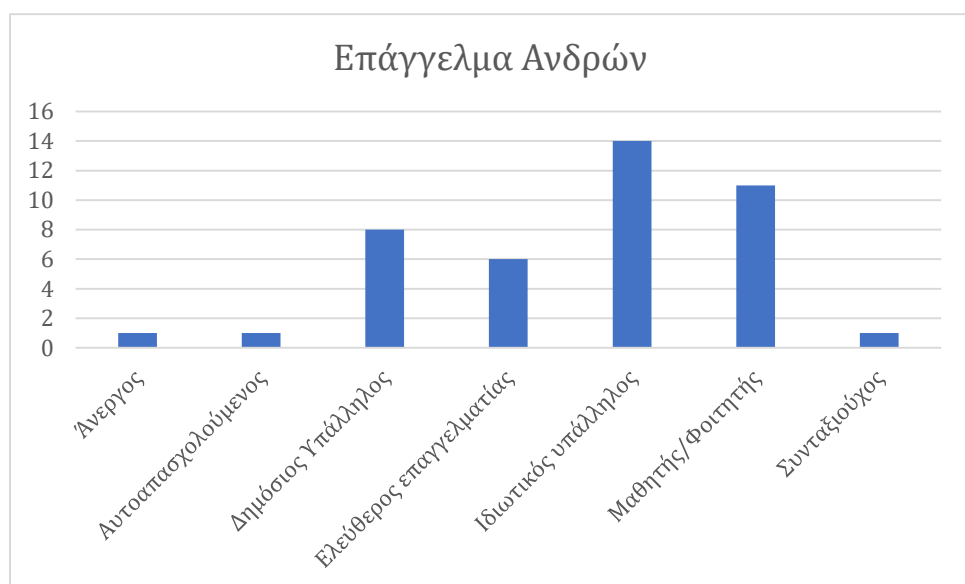


Γράφημα 4-5: Μορφωτικό επίπεδο ανδρών

Στα γραφήματα 4-4 και 4-5 εμφανίζεται το μορφωτικό επίπεδο των γυναικών και των ανδρών, αντίστοιχα. Μικρά ποσοστά του δείγματος αποτελούν οι απόφοιτοι Δημοτικού και Γυμνασίου, καθώς και οι κάτοχοι Διδακτορικού. Επομένως, το δείγμα αποτελείται κυρίως από άτομα με ανώτερη ή ανώτατη εκπαίδευση.

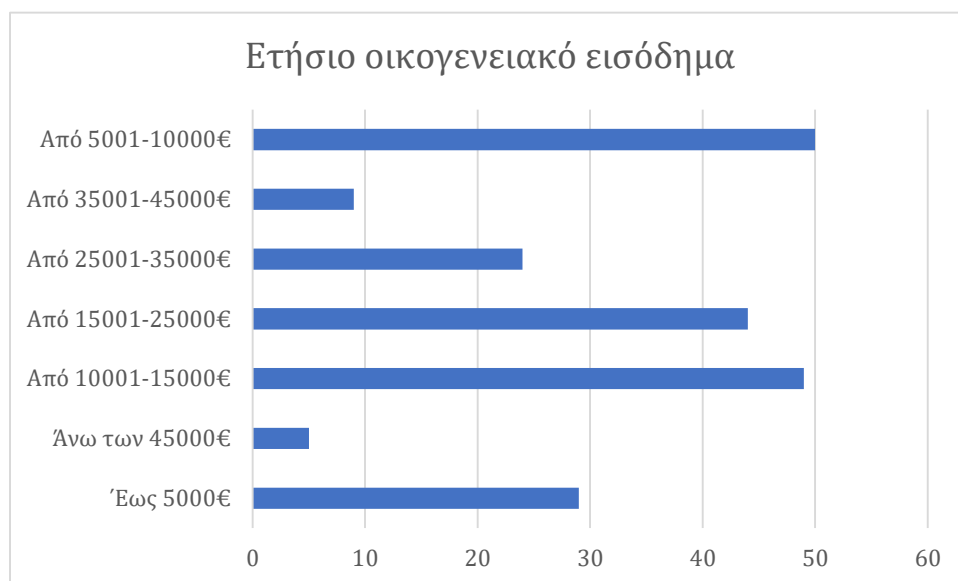


Γράφημα 4-6: Επάγγελμα Γυναικών



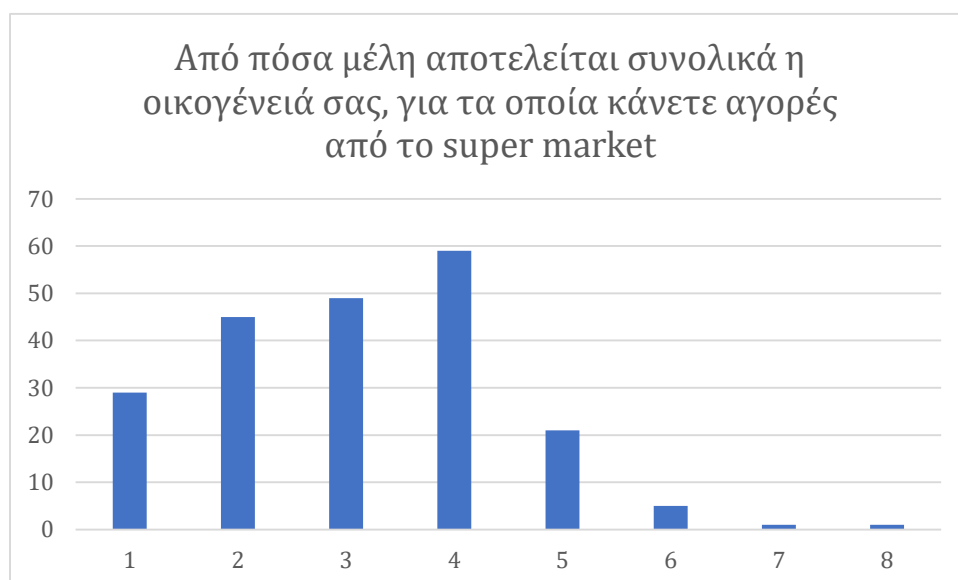
Γράφημα 4-7: Επάγγελμα Ανδρών

Στα γραφήματα 4-7 και 4-8 παρουσιάζονται τα επαγγέλματα των γυναικών και των ανδρών. Το μεγαλύτερο ποσοστό των γυναικών εργάζονται ως δημόσιοι ή ιδιωτικοί υπάλληλοι, ενώ των ανδρών επιμερίζεται σε ιδιωτικούς υπαλλήλους, φοιτητές, δημόσιους υπαλλήλους και σε ελεύθερους επαγγελματίες.



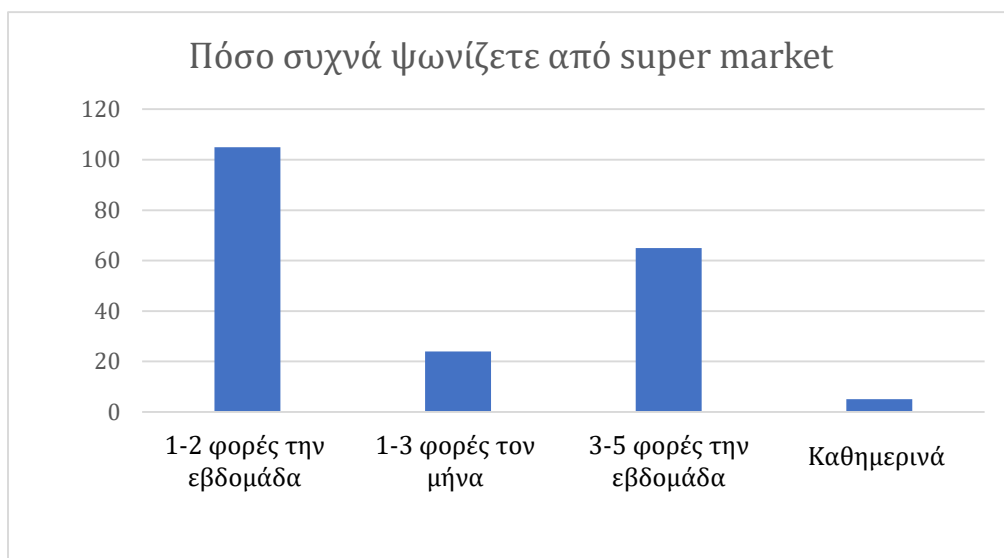
Γράφημα 4-8: Ετήσιο Οικογενειακό Εισόδημα

Στο γράφημα 4-8 παρουσιάζεται σε ποσοστά το ετήσιο εισόδημα των νοικοκυριών που συμμετείχαν στην έρευνα. Με ποσοστό 23,81% εμφανίζονται τα νοικοκυριά με εισόδημα από 5.001-10.000€, ακολουθούν με 23,33% αυτά με εισόδημα από 10001-15000€, με 20,95% τα νοικοκυριά με εισόδημα 15.001-25.000€, με 13,81% τα νοικοκυριά με εισόδημα έως 5.000€, με 11,43% αυτά με εισόδημα 25.001-35.000€ και τέλος με 4,29% και 2,38% τα νοικοκυριά με εισόδημα 35.001-45.000€ και αυτά με εισόδημα άνω των 45.000€, αντίστοιχα.



Γράφημα 4-9: Μέλη νοικοκυριού για τα οποία γίνονται αγορές από το super market

Τα μέλη του νοικοκυριού για τα οποία οι ερωτώμενοι κάνουν αγορές από το super market ως επί το πλείστον είναι 3 ή 4 και ακολουθούν όσα νοικοκυριά αποτελούνται από 1 ή 2 άτομα.



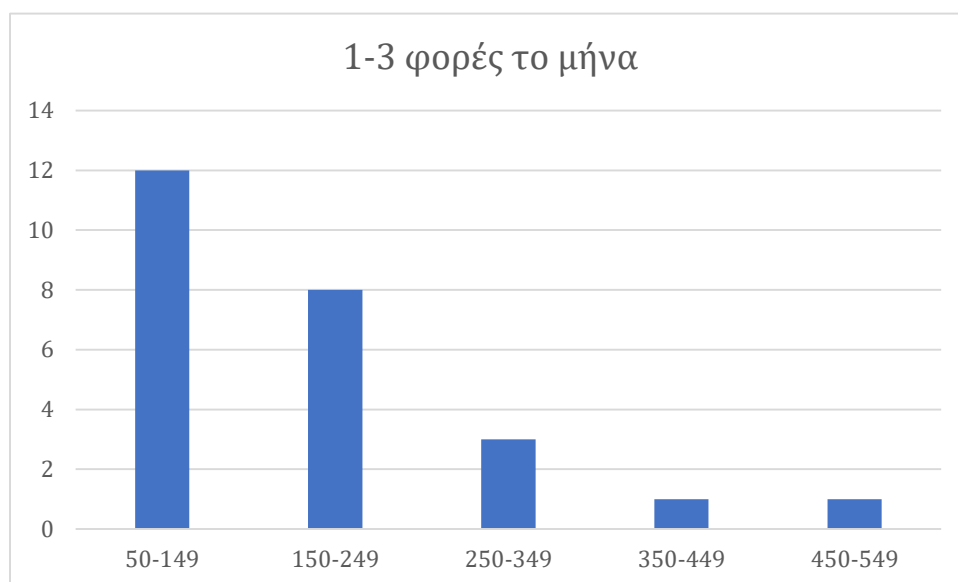
Γράφημα 4-10: Συχνότητα αγορών από super market

Στην ερώτηση για το πόσο συχνά οι καταναλωτές ψωνίζουν από super market (Γράφημα 4-10), το μεγαλύτερο ποσοστό απάντησε 1-2 φορές την εβδομάδα, αμέσως μετά ακολουθεί η απάντηση 3-5 φορές την εβδομάδα, στην συνέχεια 1-3 φορές το μήνα και τέλος καθημερινά. Από αυτήν την ερώτηση συμπεραίνεται ότι οι καταναλωτές που απάντησαν στο συγκεκριμένο ερωτηματολόγιο, έχουν μία καλή εικόνα των super market στον Ν. Χανίων, καθώς οι περισσότεροι τα επισκέπτονται πάνω από μία φορά σε εβδομαδιαία βάση.



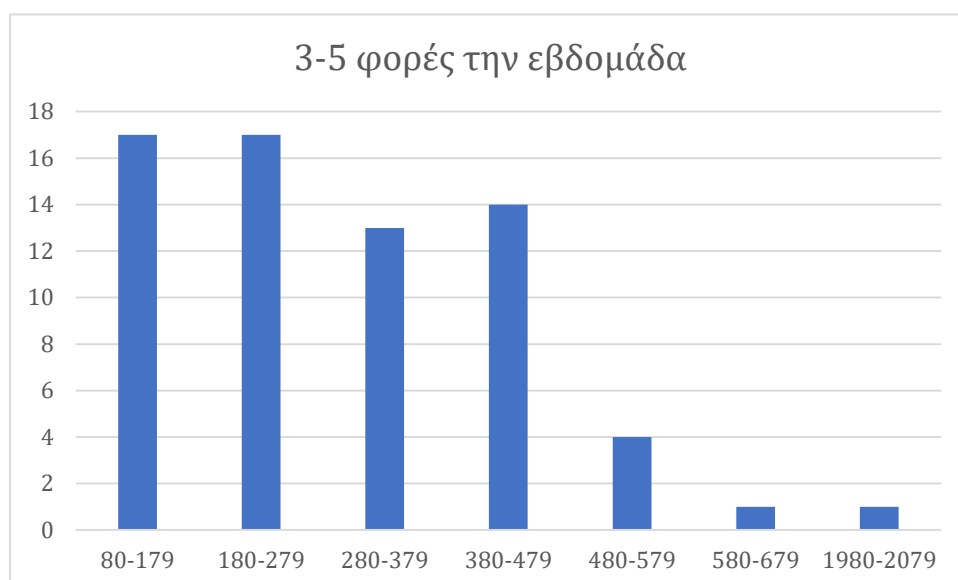
Γράφημα 4-11: Μέση μηνιαία δαπάνη καταναλωτών που επισκέπτονται το super market 1-2 φορές την εβδομάδα

Στο γράφημα 4-11 παρουσιάζεται η μέση μηνιαία δαπάνη των καταναλωτών που επισκέπτονται τα super market 1-2 φορές την εβδομάδα κατά μέσο όρο. Παρατηρείται ότι οι περισσότεροι καταναλωτές διαθέτουν από 100€ έως 199€ το μήνα καθώς και 300 έως 399€ το μήνα.



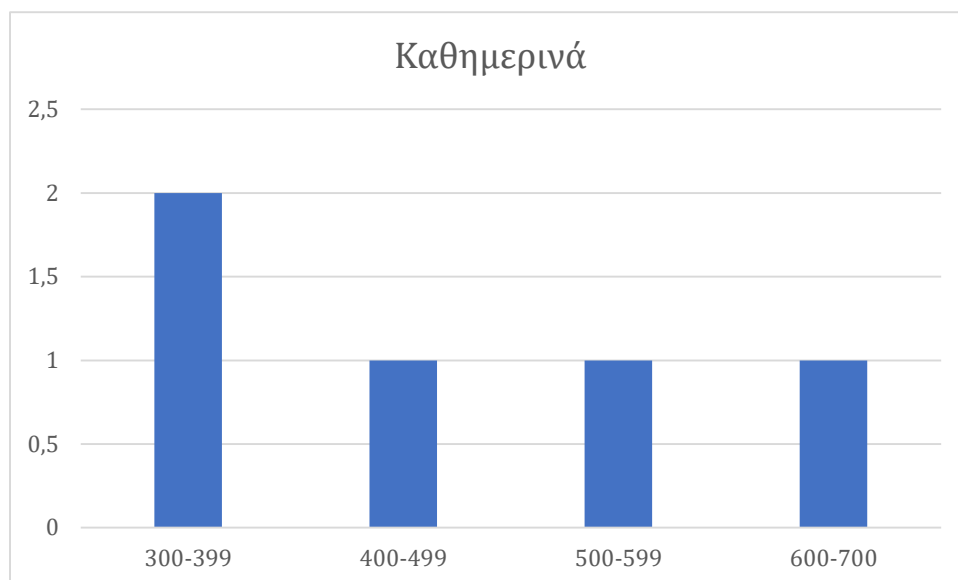
Γράφημα 4-12: Μέση μηνιαία δαπάνη καταναλωτών που επισκέπτονται το super market 1-3 φορές το μήνα

Στο γράφημα 4-12 παρουσιάζεται η μέση μηνιαία δαπάνη των καταναλωτών που επισκέπτονται τα super market 1-3 φορές το μήνα κατά μέσο όρο. Παρατηρείται ότι οι περισσότεροι καταναλωτές διαθέτουν από 50€ έως 149€ το μήνα.



Γράφημα 4-13: Μέση μηνιαία δαπάνη καταναλωτών που επισκέπτονται το super market 3-5 φορές την εβδομάδα

Στο γράφημα 4-13 παρουσιάζεται η μέση μηνιαία δαπάνη των καταναλωτών που επισκέπτονται τα super market 3-5 φορές την εβδομάδα κατά μέσο όρο. Παρατηρείται ότι οι περισσότεροι καταναλωτές διαθέτουν από 80€ έως 179€ και από 180€ έως 279€ το μήνα.



Γράφημα 4-14: Μέση μηνιαία δαπάνη καταναλωτών που επισκέπτονται το super market καθημερινά

Στο γράφημα 4-14 παρουσιάζεται η μέση μηνιαία δαπάνη των καταναλωτών που επισκέπτονται τα super market καθημερινά. Παρατηρείται ότι οι περισσότεροι καταναλωτές διαθέτουν από 300€ έως 399€ το μήνα.

Συμπεράσματα Περιγραφικής Στατιστικής

Από τα αποτελέσματα της στατιστικής ανάλυσης συμπεραίνεται ότι αρχικά το μεγαλύτερο ποσοστό των ερωτηθέντων αφορά σε γυναίκες 79%, ενώ το ποσοστό των ανδρών είναι 21%. Η μεγαλύτερη ηλικιακή ομάδα για τις γυναίκες είναι αυτή των 30-39 ετών, ενώ για τους άνδρες 20-29 ετών. Όσον αφορά στο μορφωτικό επίπεδο των ερωτηθέντων, τόσο για τις γυναίκες, όσο και για τους άνδρες, η μεγαλύτερη ομάδα είναι πτυχιούχοι τριτοβάθμιας εκπαίδευσης. Το ίδιο ισχύει και για το επάγγελμα των ερωτηθέντων, όπου και για τα δύο φύλα η μεγαλύτερη ομάδα αφορά ιδιωτικούς υπαλλήλους. Το μεγαλύτερο σε ποσοστό ετήσιο οικογενειακό εισόδημα των νοικοκυριών που απάντησαν στο ερωτηματολόγιο είναι από 5001€- 10000€ και οι περισσότερες οικογένειες αποτελούνται από 4 μέλη. Οι περισσότεροι ερωτώμενοι πραγματοποιούν αγορές από super market 1-2 φορές την εβδομάδα. Όσον αφορά την μέση μηνιαία δαπάνη όσων πραγματοποιούν αγορές από super market συμπεραίνεται πως αυτοί που επισκέπτονται τα super market 1-2 φορές την εβδομάδα δαπανούν μηνιαίως 100-199€, όσοι τα επισκέπτονται 1-3 φορές τον μήνα δαπανούν οι περισσότεροι 50-149€, αυτοί που τα επισκέπτονται 3-5 φορές την εβδομάδα δαπανούν

οι περισσότεροι 80-179€ μηνιαίως, ενώ όσοι τα επισκέπτονται καθημερινά δαπανούν 300-399€ μηνιαίως.

4.3. ΠΡΟΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ

4.3.1. Εισαγωγή δεδομένων στο RapidMiner

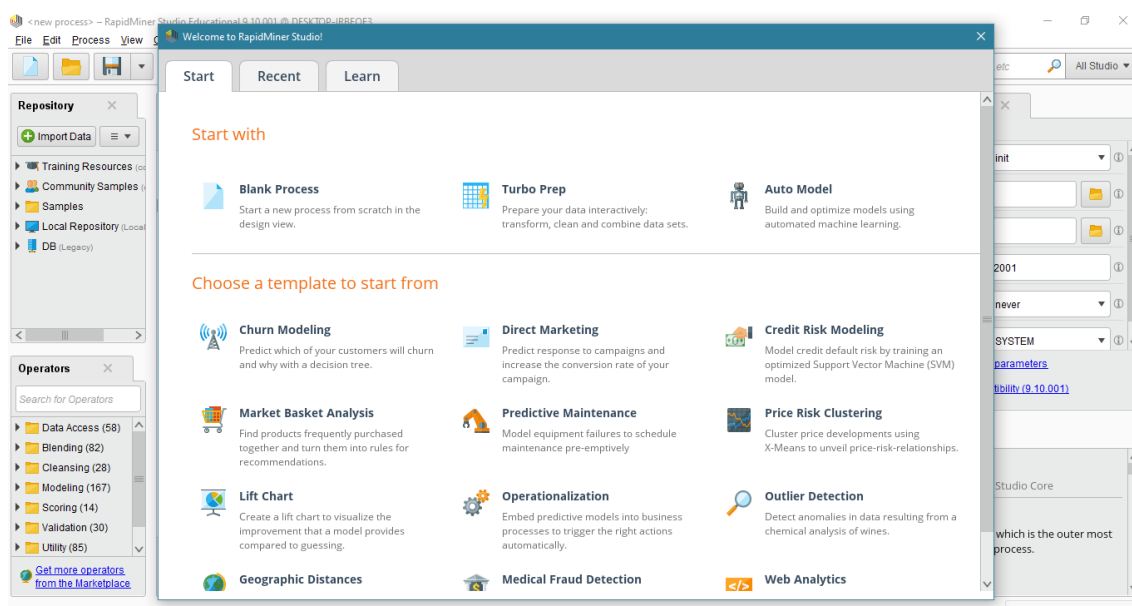
Η προεπεξεργασία δεδομένων είναι μία διαδικασία απαραίτητη προκειμένου να μην εξαχθούν λανθασμένα αποτελέσματα από τις αναλύσεις που θα ακολουθήσουν (βλ. παρ. 3.4).

Η προεπεξεργασία των δεδομένων θα γίνει στο λογισμικό RapidMiner.

Το RapidMiner (<https://rapidminer.com/>) είναι ένα λογισμικό το οποίο χρησιμοποιείται στον τομέα της ανάλυσης δεδομένων και είναι σχεδιασμένο για επιχειρήσεις. Το συγκεκριμένο λογισμικό παρέχει μία προηγμένη λύση μειώνοντας τα σφάλματα, σχεδόν εξαλείφοντας την ανάγκη γραφής κώδικα και γι' αυτό το λόγο είναι πολύ φιλικό προς τον οποιοδήποτε χρήστη. Παρέχει διαδικασίες εξόρυξης δεδομένων και μηχανικής μάθησης, όπως: φόρτωση και μετασχηματισμός δεδομένων, προεπεξεργασία και οπτικοποίηση δεδομένων, προγνωστική ανάλυση και στατιστική μοντελοποίηση, αξιολόγηση και ανάπτυξη και είναι γραμμένο σε γλώσσα προγραμματισμού Java.

Αρχικά, τα δεδομένα αποθηκεύονται σε μορφή .xlsx και στη συνέχεια πρέπει να εισαχθούν στο RapidMiner για την προεπεξεργασία τους.

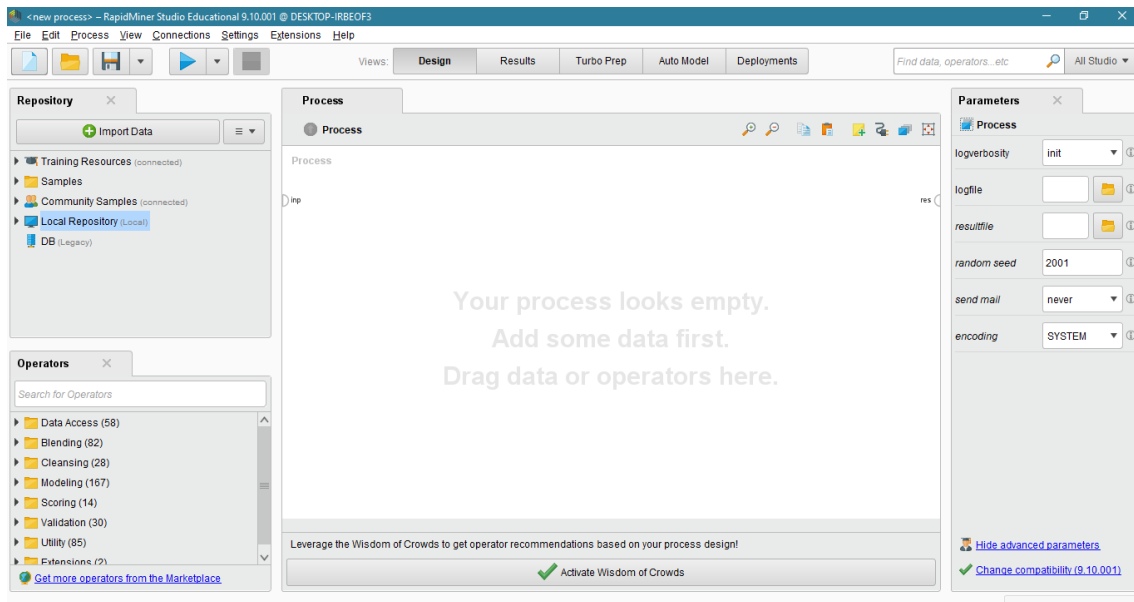
Η αρχική οθόνη του RapidMiner φαίνεται στην εικόνα 4-1.



Εικόνα 4-1: Αρχική οθόνη RapidMiner

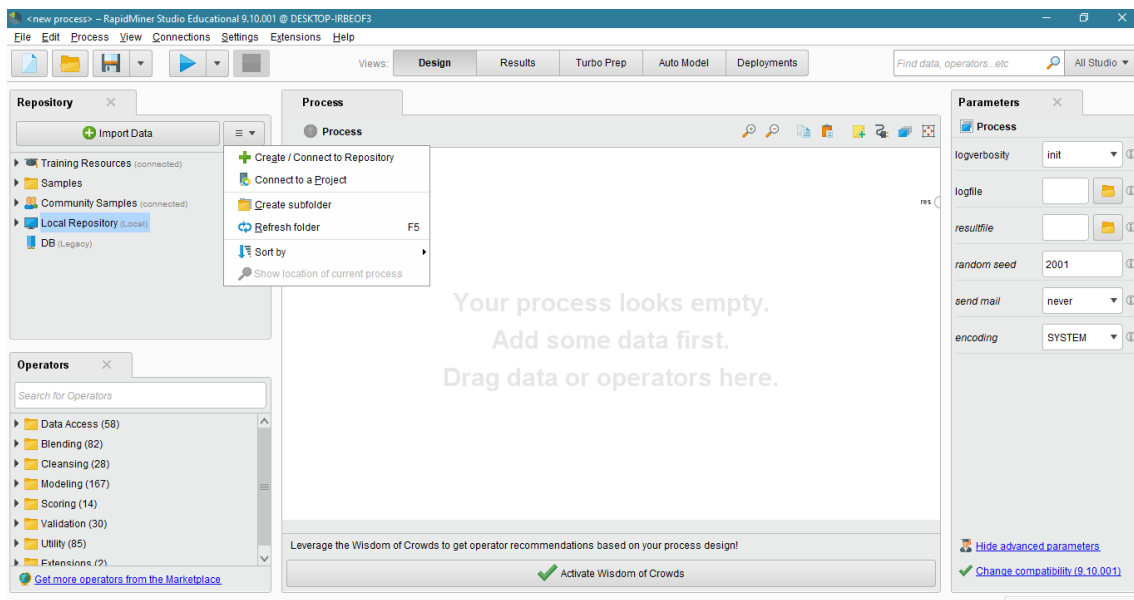
Στην συνέχεια, πρέπει να εισαχθούν τα δεδομένα. Τα δεδομένα αρχικά εισάγονται στη βάση δεδομένων του λογισμικού και στην συνέχεια πραγματοποιείται η προεπεξεργασία.

Επομένως, στο παράθυρο που εμφανίζεται πρώτα, επιλέγεται η εντολή “Blank Process”. Η οθόνη που εμφανίζεται φαίνεται στην εικόνα 4-2.



Εικόνα 4-2: Αρχική οθόνη

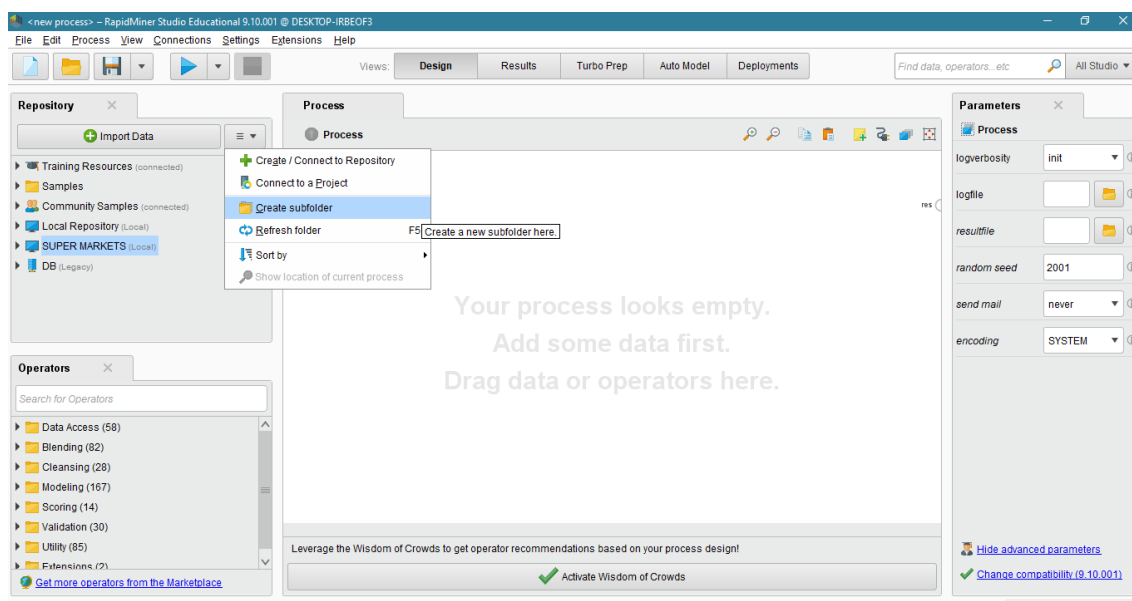
Έπειτα, πρέπει να δημιουργηθεί στο συγκεκριμένο σημείο ένα αποθετήριο. Αυτό δημιουργείται πατώντας το πλήκτρο δίπλα στο “Import data”.



Εικόνα 4-3: Δημιουργία Αποθετηρίου

Ο χρήστης στο συγκεκριμένο σημείο επιλέγει “Create/Connect to Repository”. Στο αμέσως επόμενο βήμα πρέπει να δώσει όνομα στο αποθετήριο που θα δημιουργήσει καθώς και το σημείο που θέλει να αποθηκευτεί.

Στη συνέχεια, δημιουργεί δύο υποφακέλους. Έναν για τα δεδομένα του και έναν για να αποθηκεύσει τις διάφορες διαδικασίες επεξεργασίας των δεδομένων που πρόκειται να υπολοιστεί. (Εικόνα 4-4)



Εικόνα 4-4: Εισαγωγή υποφακέλων στο αποθετήριο

Πλέον, ο χρήστης είναι έτοιμος για να ξεκινήσει την εισαγωγή των δεδομένων.

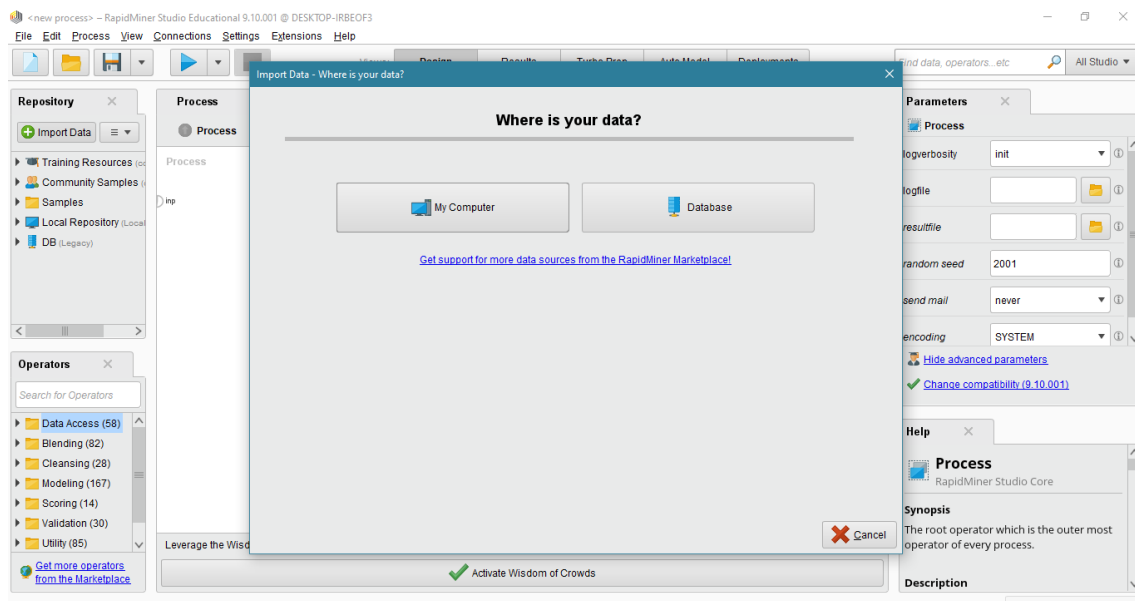
Αρχικά επιλέγει την εντολή “Import data” που βρίσκεται πάνω αριστερά της οθόνης.

Τα βήματα που πρέπει να ακολουθήσει ο χρήστης στη συνέχεια, προκειμένου να εισάγει τα δεδομένα είναι τα εξής:

Βήμα 1^ο: Επιλέγει από πού πρέπει να αντληθούν τα δεδομένα (Εικόνα 4-5).

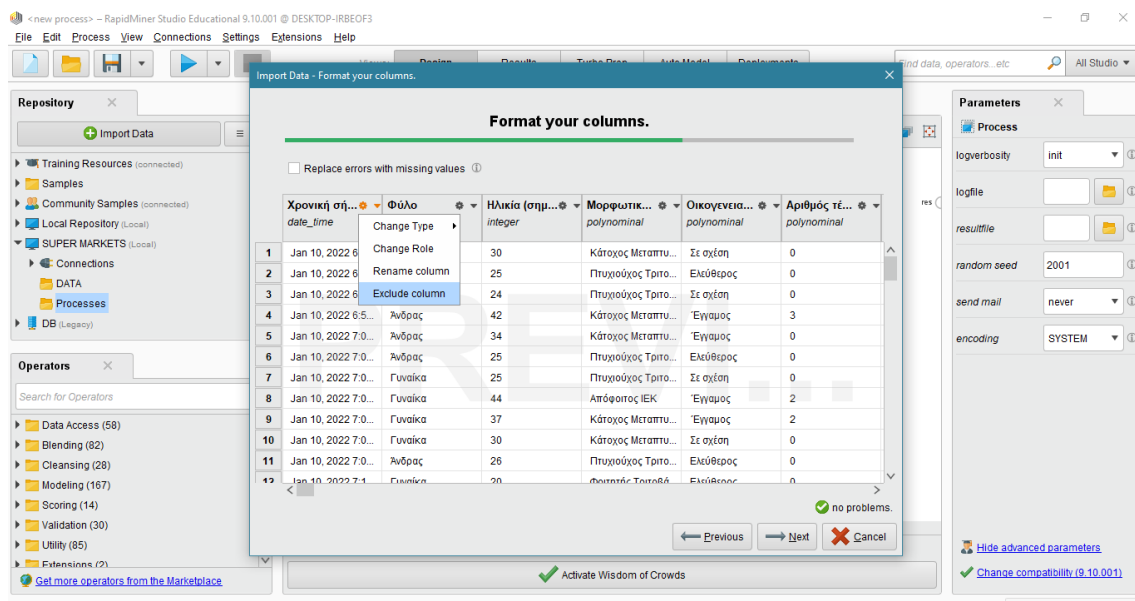
Βήμα 2^ο: Στο δεύτερο βήμα επιλέγει ποιες στήλες από το αρχείο θέλει να εισαχθούν.

Βήμα 3^ο: Επιλέγει τη μορφοποίηση του αρχείου. Ο χρήστης δηλαδή έχει την δυνατότητα να επιλέξει ανα στήλη αν θέλει να της αλλάξει τον τύπο δεδομένων, να την μετονομάσει ή να την αποκλείσει.



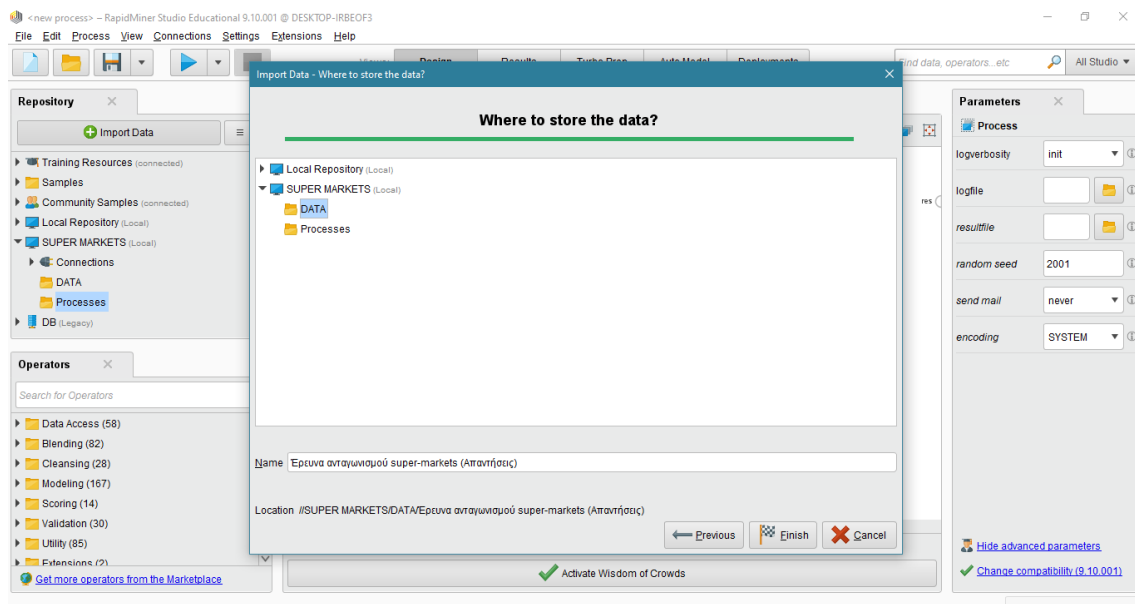
Εικόνα 4-5: Εισαγωγή Δεδομένων (Βήμα 1^ο)

Στο συγκεκριμένο βήμα θα αποκλειστεί η στήλη «Χρονική σήμανση», καθώς περιέχει ημερομηνία και ώρα συμπλήρωσης του κάθε ερωτηματολογίου, πράγμα που δεν είναι απαραίτητο στην μετέπειτα επεξεργασία. (Εικόνα 4-6)



Εικόνα 4-6: Επεξεργασία στηλών πριν την αποθήκευση

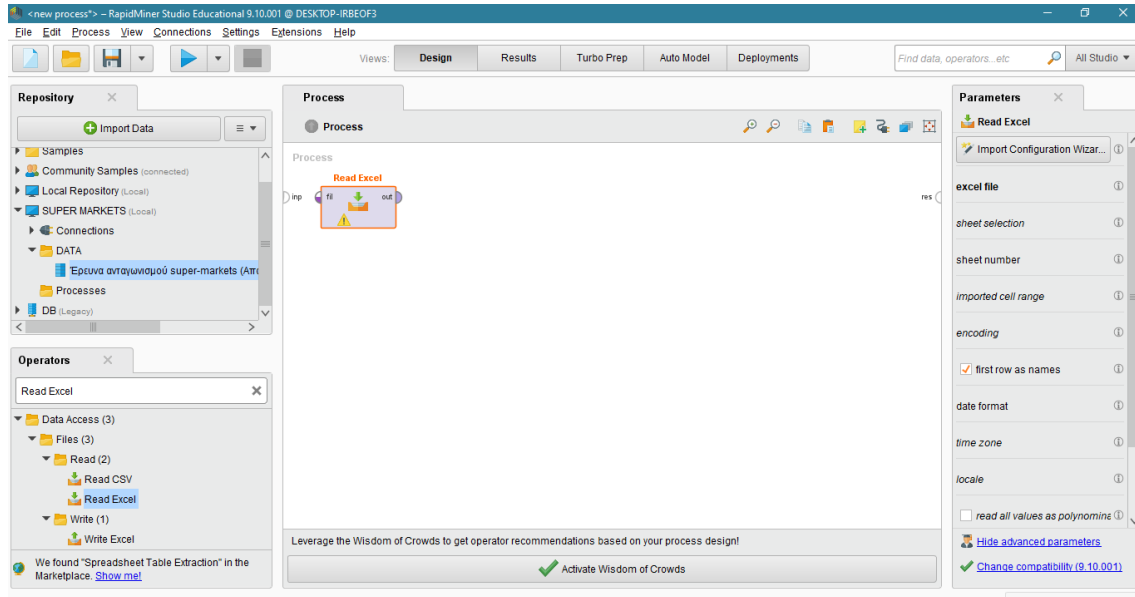
Βήμα 4^ο: Επιλέγει πού θέλει να αποθηκευτούν τα δεδομένα. (Εικόνα 4-7)



Εικόνα 4-7: Αποθήκευση δεδομένων

4.3.2. Προεπεξεργασία Δεδομένων

Στο επόμενο βήμα θα ξεκινήσει η προεπεξεργασία των δεδομένων. Στην καρτέλα “Design” επιλέγεται από τους Operators, ο “Read Excel”, όπως φαίνεται στην εικόνα 4-8.



Εικόνα 4-8: Operator “Read Excel”

Στη συνέχεια εισάγονται τα δεδομένα στον operator από την καρτέλα Parameters-> Import Configuration Wizard. Ενώνεται ο operator με το “res” (results) στην κεντρική καρτέλα και εκτελείται η διαδικασία.

Όταν η διαδικασία αυτή ολοκληρωθεί, τα δεδομένα θα εμφανιστούν στην οθόνη. Αριστερά εμφανίζονται καρτέλες με διάφορες λειτουργίες. Ο χρήστης πατώντας την καρτέλα “Statistics” μπορεί να εντοπίσει τις ελλιπείς τιμές από κάθε στήλη των δεδομένων του, τις ελάχιστες και μέγιστες τιμές, καθώς και όλες τις εναλλακτικές επιλογές που δόθηκαν από τους ερωτηθέντες.

4.3.2.1. Εύρεση και αντικατάσταση ελλιπών τιμών

Στα δεδομένα που συλλέχθηκαν με χρήση ερωτηματολογίου εμφανίστηκαν ελλιπείς τιμές: 1 ελλιπής τιμή στην ερώτηση «Φύλο», 52 ελλιπείς τιμές στην ερώτηση «Μπορείτε να μας αναφέρετε αν εκτός από τα παραπάνω super markets, αγοράζετε προϊόντα από άλλα καταστήματα; Ποια και γιατί;» και 166 στην ερώτηση «Παρατηρήσεις / σχόλια / προτάσεις που θέλετε να προσθέσετε για τα Super Markets της περιοχής σας». Οι ελλιπείς τιμές είναι λίγες, καθώς το ηλεκτρονικό ερωτηματολόγιο δεν επιτρέπει στο χρήστη να παραβλέψει κάποιες ερωτήσεις.

Υπάρχουν, ωστόσο, και τιμές οι οποίες δεν είναι εντός των αναμενόμενων, όπως η απάντηση στην ερώτηση «Αριθμός τέκνων» ήταν «Σύντομα 1», καθώς και απαντήσεις εκτός των προτεινόμενων για την ερώτηση «Επάγγελμα». Όλες οι ελλιπείς τιμές και οι διαφορετικές απαντήσεις που έχουν δοθεί φαίνονται συγκεντρωμένες και είναι εύκολα εντοπίσιμες στην καρτέλα results μετά την εισαγωγή των δεδομένων. Η οθόνη αυτή εμφανίζεται στην εικόνα 4-9.

Name	Type	Missing	Statistics
Φύλο	Nominal	1	Least: Άνδρας (43), Most: Γυναίκα (166), Values: Γυναίκα (166), Άνδρας (43)
Ηλικία (σημειώστε την απάντη...	Integer	0	Min: 20, Max: 76, Average: 37.181
Μορφωτικό επίπεδο	Nominal	0	Least: Αποφοιτος Δει (1), Most: Πτυχιούχ [...] υσης (72), Values: Πτυχιούχ [...] παιδείας (72), Απόφοιτος Λυκείου (40), ...[7]
Οικογενειακή κατάσταση	Nominal	0	Least: Σε σχέση (41), Most: Έγγαμος (123), Values: Έγγαμος (123), Ελεύθερος (46), ...[1 more]
Αριθμός τέκνων (σημειώστε τη...	Nominal	0	Least: Σύντομα 1 (1), Most: 0 (77), Values: 0 (77), 2 (63), ...[5 more]
Από πόσα μέλη αποτελείται συ...	Integer	0	Min: 1, Max: 9, Average: 3.110
Επάγγελμα	Nominal	0	Least: μαγειρίσσα (1), Most: Ιδιωτικός υπάλληλος (72), Values: Ιδιωτικός υπάλληλος (72), Δημόσιος Υπάλληλος (42), ...[11]

Εικόνα 4-9: Εμφάνιση Δεδομένων συγκεντρωτικά

Στη συνέχεια, επιλέγεται αρχικά η αντικατάσταση των ελλιπων τιμών.

- Στο παράθυρο των Operators επιλέγεται ο operator “Replace missing values” και τοποθετείται μετά τον operator “Read excel” στο κεντρικό παράθυρο του Design.
- Στο παράθυρο των Parameters, επιλέγεται για ποιο χαρακτηριστικό θα γίνει αντικατάσταση ελλιπών τιμών. Στη συγκεκριμένη περίπτωση στο πεδίο “attribute filter type” δίνεται η επιλογή “Single” καθώς αφορά σε ένα μόνο χαρακτηριστικό η διαδικασία. Στο πεδίο “attribute” επιλέγεται το «Φύλο» και στη συνέχεια στο πεδίο “default”, η επιλογή value. Στην συνέχεια εισάγεται η τιμή «Γυναίκα», καθώς είναι η επιλογή με το μεγαλύτερο πλήθος απαντήσεων και αφορά κατηγορική μεταβλητή, επομένως είναι αδύνατον να βρεθεί μέσος όρος.

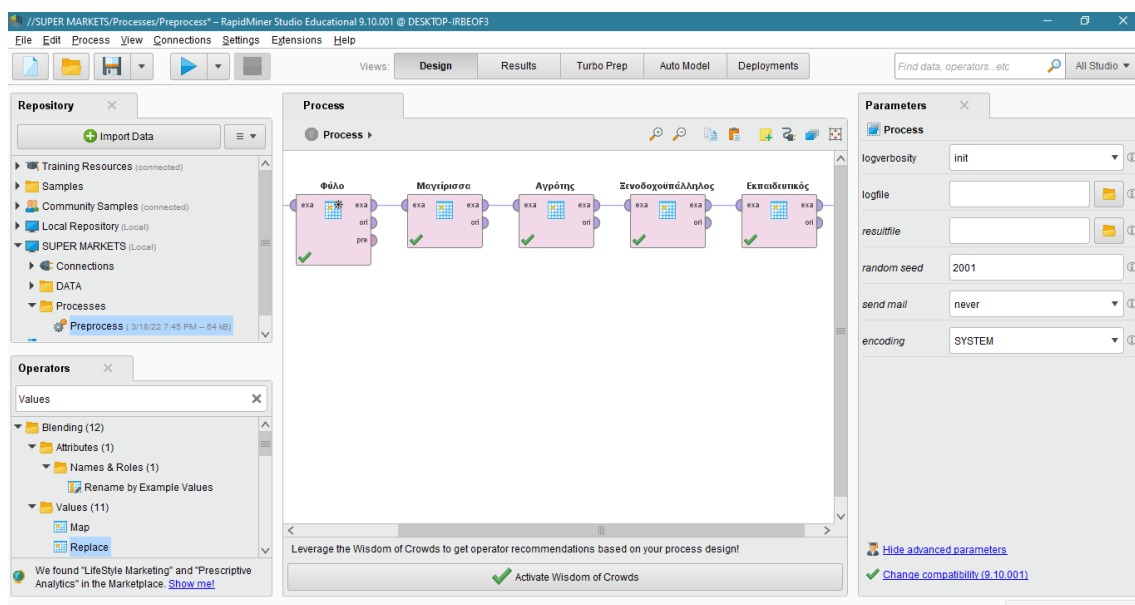
Μετά την εκτέλεση του παραπάνω operator, ελέγχονται ξανά οι τιμές που έχουν δοθεί στις απαντήσεις από την καρτέλα “Statistics”.

Παρατηρείται ότι υπάρχει πρόβλημα στις ερωτήσεις:

- Αριθμός τέκνων,
- Μόρφωση,
- Επάγγελμα και
- Περιοχή

Για την διόρθωση των παραπάνω attributes, θα χρησιμοποιηθεί ο operator “replace” από την καρτέλα των operators “values”. Στο πεδίο “attribute filter type” δίνεται η επιλογή “Single” καθώς αφορά σε ένα μόνο χαρακτηριστικό η διαδικασία και στο πεδίο “attribute” δίνεται η τιμή «Επάγγελμα» αρχικά. Στη συνέχεια, εισάγεται η λανθασμένη τιμή στο πεδίο “replace what” και στο πεδίο “replace by” εισάγεται η σωστή τιμή. Για κάθε λανθασμένη τιμή που έχει καταγραφεί προστίθεται ένας καινούριος operator με τον ίδιο τρόπο. Το ίδιο θα συμβεί και στα υπόλοιπα χαρακτηριστικά.

Οι operators εμφανίζονται στην κεντρική καρτέλα του Design όπως στην εικόνα 4-10. Εφαρμόζονται πολλοί operators, και ειδικότερα ένας για κάθε διαφορετική απάντηση, η οποία μπορεί να ενσωματωθεί στις ήδη υπάρχουσες εναλλακτικές που δίνονται.



Εικόνα 4-10: Εμφάνιση operators

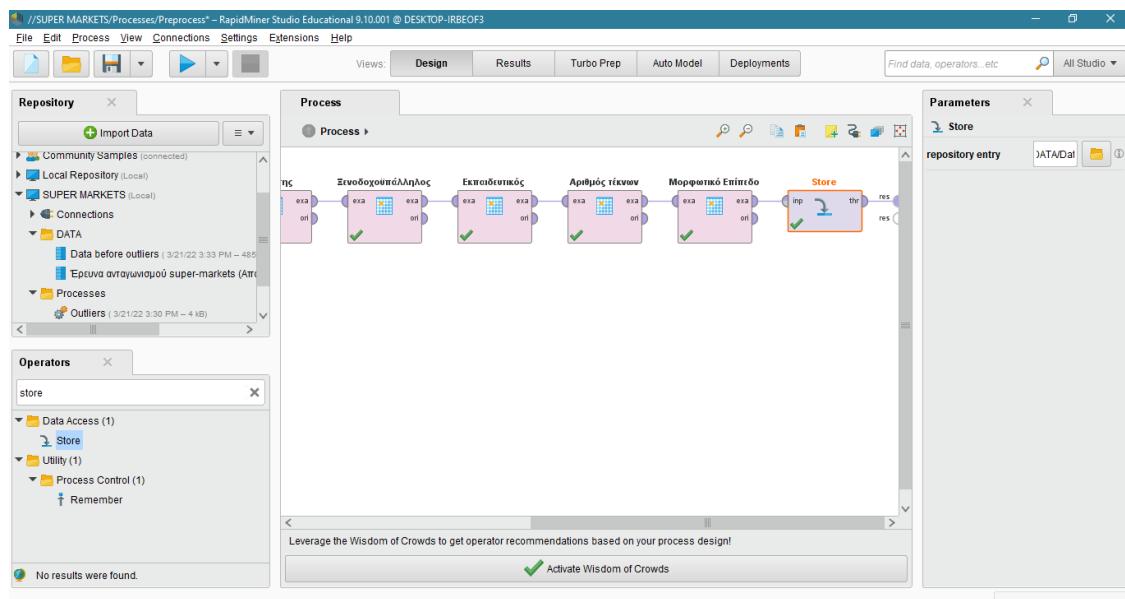
Οι περιοχές αλλάχτηκαν στο αρχικό αρχείο δεδομένων, καθώς ήταν ερώτηση ανοιχτού τύπου και οι απαντήσεις που δόθηκαν, ενώ ήταν έγκυρες δεν μπορούσαν να ομαδοποιηθούν λόγω ορθογραφικών λαθών, εισαγωγή κενού μετά την απάντηση, χρήση λατινικού αλφαβήτου και έλλειψης τονισμού. Επομένως, επειδή αυτή η διαδικασία για καθεμία απάντηση είναι χρονοβόρα, επιλέχθηκε η διόρθωση μέσω Excel.

Συγκεκριμένα, οι απαντήσεις οι οποίες άλλαξαν είναι οι εξής:

Πίνακας 4-1: Διορθωμένες απαντήσεις

Χαρακτηριστικό	Απάντηση
Επάγγελμα	Μαγειρίσσια, Εκπαιδευτικός, Αγρότης, Ξενοδοχοϋπάλληλος
Αριθμός τέκνων	Σύντομα 1
Μορφωτικό επίπεδο	Απόφοιτος Αει
Περιοχές	Διάφορες

Για την αποθήκευση των απαντήσεων και για την χρήση τους σε επόμενη διαδικασία, θα χρησιμοποιηθεί ο operator "Store", ο οποίος τοποθετείται τελευταίος στην γραμμή των operators, όπως στην εικόνα 4-11.



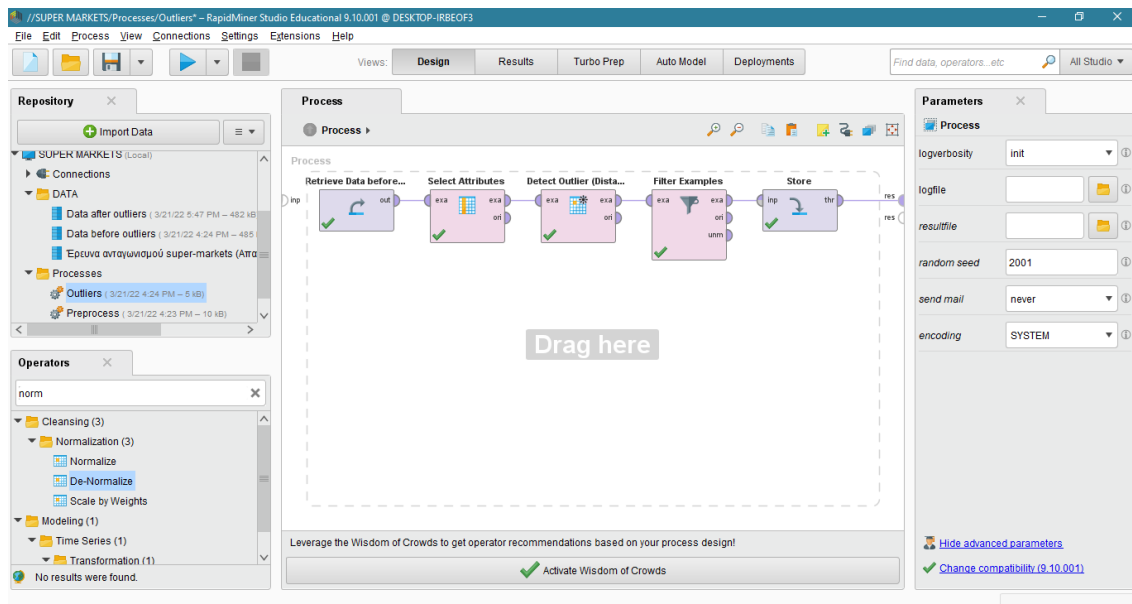
Εικόνα 4-11: Αποθήκευση δεδομένων μετά την πρώτη επεξεργασία

4.3.2.2. Εντοπισμός και απομάκρυνση ακραίων τιμών

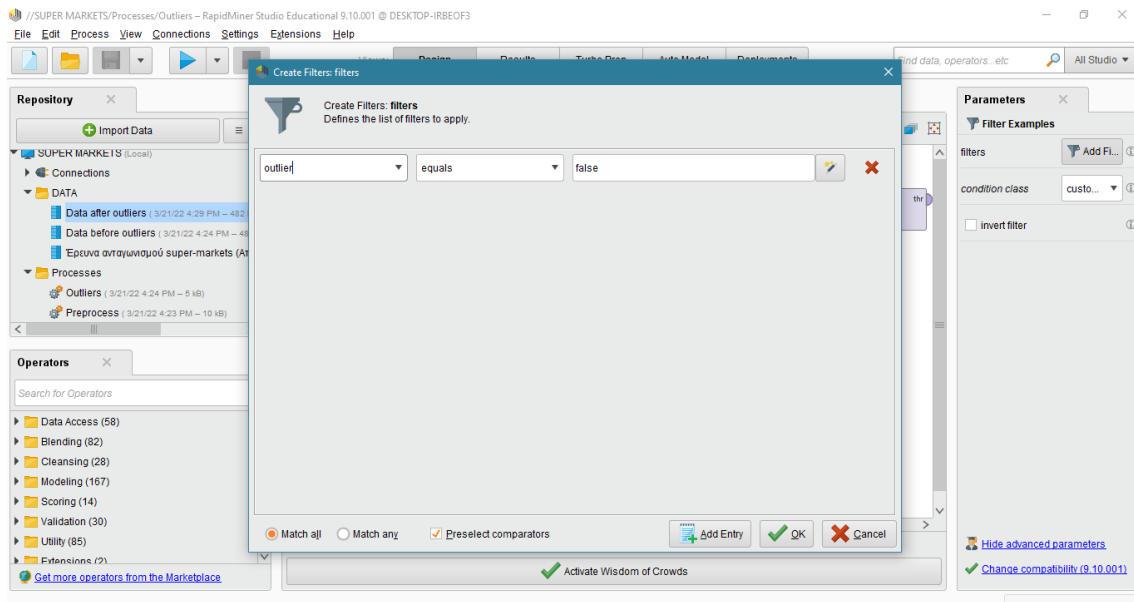
Για τον πλήρη καθαρισμό των δεδομένων απομένει ο εντοπισμός τυχών ακραίων τιμών. Τα δεδομένα που αποθηκεύτηκαν από την προηγούμενη επεξεργασία θα χρησιμοποιηθούν στην νέα. Επομένως, δημιουργείται “New process”-> “Blank process”.

Αρχικά, εισάγονται τα δεδομένα. Στη συνέχεια, εισάγονται στη σειρά οι operators “Select attributes”, “Detect Outlier (Distances)”, “Filter examples” και “Store”. Τα δεδομένα θα υποβληθούν σε κανονικοποίηση προκειμένου να γίνει δυνατός ο εντοπισμός των ακραίων τιμών. Οι παράμετροι σε όλους τους operators, εκτός από τον “Filter examples” κρατάνε τις τιμές που έχουν by default.

Στον operator “Filter examples”, στις παραμέτρους (“Parameters”) επιλέγεται το “outlier”, στη συνέχεια το “equals to” και τέλος ο χρήστης επιλέγει “false”. Αυτό σημαίνει ότι κάθε ακραία τιμή παίρνει την τιμή True και έπειτα διαγράφεται. Στις εικόνες 4-12 και 4-13, φαίνεται η διαδικασία που περιγράφεται παραπάνω, ενώ στην εικόνα 4-14 το αποτέλεσμα της διαδικασίας.



Εικόνα 4-12: Διαδικασία εύρεσης ακράιων τιμών



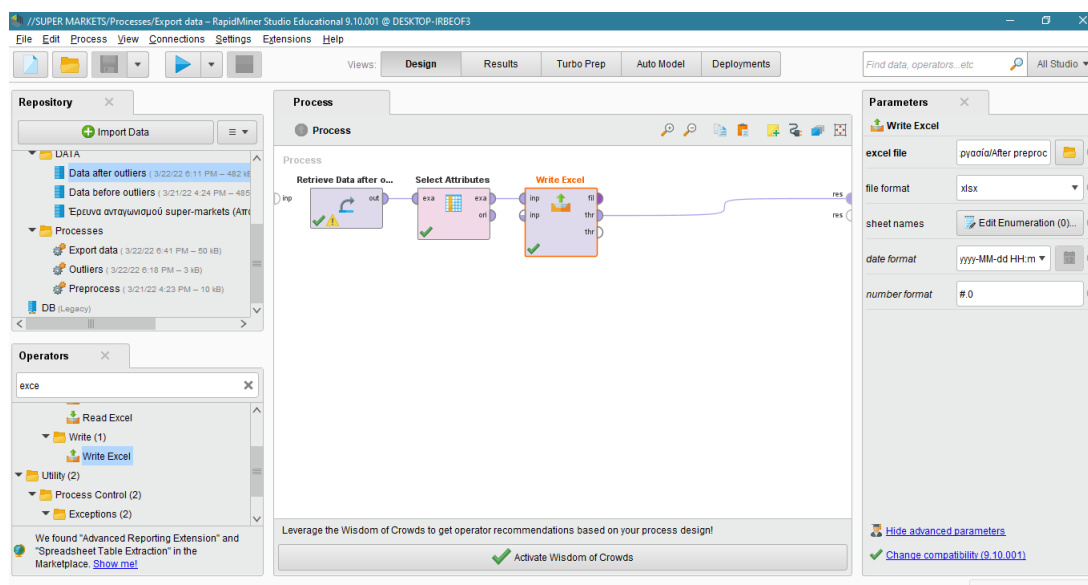
Εικόνα 4-13: Ορισμός παραμέτρων "Filter examples"

Row No.	outlier	Μορφωτικό ...	Αριθμός τέκν...	Επάγγελμα	Φύλο	Χρονική σήμ...	Ηλικία (σημ...	Οικογενειακ...	Από πόσα μ...	Ετήσιο οικογ...	Θα σας παρ...	Π
40	false	Φοιτητής Τριτ...	0	Μαθητής/Φοιτ...	Γυναίκα	Jan 16, 2022 ...	30	Σε σχέση	1	Από 5001-10...	Αμπεριά	1
41	false	Φοιτητής Τριτ...	0	Μαθητής/Φοιτ...	Ανδρας	Jan 16, 2022 ...	23	Ελεύθερος	1	Έως 5000€	Ακρωτήρι	1
42	false	Φοιτητής Τριτ...	0	Μαθητής/Φοιτ...	Γυναίκα	Jan 16, 2022 ...	23	Ελεύθερος	4	Από 15001-2...	Κέντρο	1
43	false	Πτυχιούχος Τ...	2	Δημόσιος Υπ...	Γυναίκα	Jan 16, 2022 ...	38	Έγγαμος	4	Από 25001-3...	Βουκολιές	1
44	false	Φοιτητής Τριτ...	0	Μαθητής/Φοιτ...	Ανδρας	Jan 16, 2022 ...	22	Σε σχέση	1	Έως 5000€	Ακρωτήρι	1
45	false	Πτυχιούχος Τ...	3	Ιδιωτικός υπά...	Γυναίκα	Jan 16, 2022 ...	38	Έγγαμος	5	Από 15001-2...	Ακρωτήρι	3
46	false	Απόφοιτος Λυ...	3	Αυτοαπασχολ...	Γυναίκα	Jan 16, 2022 ...	34	Έγγαμος	5	Από 5001-10...	Κίσαμος	1
47	false	Κάτοχος Μετα...	2	Ιδιωτικός υπά...	Γυναίκα	Jan 16, 2022 ...	40	Έγγαμος	4	Από 5001-10...	Μουραίες	1
48	false	Απόφοιτος ΙΕΚ	2	Δημόσιος Υπ...	Γυναίκα	Jan 16, 2022 ...	40	Έγγαμος	4	Από 25001-3...	Κέντρο	1
49	false	Κάτοχος Μετα...	0	Άνεργος	Γυναίκα	Jan 16, 2022 ...	28	Σε σχέση	2	Από 5001-10...	Σούδα	1
50	false	Πτυχιούχος Τ...	0	Ιδιωτικός υπά...	Γυναίκα	Jan 16, 2022 ...	37	Έγγαμος	2	Από 10001-1...	Βαρόπτερο	1
51	false	Απόφοιτος ΙΕΚ	2	Ιδιωτικός υπά...	Γυναίκα	Jan 16, 2022 ...	36	Έγγαμος	4	Από 10001-1...	Βαρόπτερο	3
52	false	Πτυχιούχος Τ...	3	Ιδιωτικός υπά...	Γυναίκα	Jan 16, 2022 ...	41	Έγγαμος	5	Από 25001-3...	Χαλέπα	1
53	false	Πτυχιούχος Τ...	1	Ιδιωτικός υπά...	Γυναίκα	Jan 16, 2022 ...	41	Ελεύθερος	2	Από 5001-10...	Ακρωτήρι	1

Εικόνα 4-14: Αποτελέσματα διαδικασίας εύρεσης ακραίων τιμών

Μετά το τέλος αυτής της διαδικασίας έχουν απομείνει 199 τιμές για τελική επεξεργασία, καθώς 11 εγγραφές αποτελούσαν ακραίες τιμές και αφαιρέθηκαν.

Προκειμένου τα δεδομένα να εξαχθούν και να αποθηκευτούν ξανά σε μορφή .xlsx, για να υποβληθούν στις επόμενες επεξεργασίες, δημιουργήθηκε μία νέα επεξεργασία. Σε αυτήν την επεξεργασία χρησιμοποιήθηκε το τελευταίο dataset των δεδομένων (μετά την ολοκλήρωση του καθαρισμού των δεδομένων) και στη συνέχεια ο operator “select attributes” προκειμένου να χρησιμοποιηθούν μόνο τα δεδομένα που χρειάζονται. Αφέθηκαν, δηλαδή, εκτός τα χαρακτηριστικά “outliers” και “χρονική σήμανση”. Στο τέλος της συγκεκριμένης διαδικασίας χρησιμοποιήθηκε ο operator “Write Excel”, στις παραμέτρους του οποίου, επιλέχθηκε η τοποθεσία αποθήκευσης των δεδομένων προς εξαγωγή. (Εικόνα 4-15)



Εικόνα 4-15: Εξαγωγή Δεδομένων σε μορφή .xlsx

Συμπεράσματα προεπεξεργασίας δεδομένων

Από τα αποτελέσματα της προεπεξεργασίας δεδομένων προκύπτει ότι μετά την εύρεση και αντικατάσταση ελλιπών τιμών και ακραίων τιμών οι εγγραφές που απέμειναν ήταν 199.

4.4. ΠΟΛΥΚΡΙΤΗΡΙΑ ΑΝΑΛΥΣΗ UTASTAR

4.4.1. Εισαγωγή

Προκειμένου να αναλυθεί ο πολυκριτήριος πίνακας, ο οποίος αποτελείται από 21 κριτήρια και 5 εναλλακτικές, είναι απαραίτητη η κατάλληλη μορφοποίησή του και η εισαγωγή αυτών των δεδομένων στο MARKEK. Έτσι, δημιουργήθηκε ένα λογιστικό φύλλο στο Excel προκειμένου να απομονωθούν τα συγκεκριμένα δεδομένα, να μορφοποιηθούν κατάλληλα, να εισαχθούν στο MARKEK και κατ' επέκταση να εφαρμοστεί η UTASTAR. Ενδεικτικά παρατίθεται η παρακάτω εικόνα (Εικόνα 4-16) με το αρχείο εισόδου που χρησιμοποιήθηκε για το MARKEK. Συγκεκριμένα το φύλλο excel αποτελείται από τέσσερα φύλλα, όπως φαίνεται στην Εικόνα 4-16. Στο πρώτο φύλλο, Options, καταχωρούνται οι πληροφορίες της έρευνας. Στο δεύτερο φύλλο, Criteria, καταχωρούνται οι λεπτομέρειες που αφορούν στα κριτήρια. Στο τρίτο φύλλο, Alternatives names, καταχωρούνται οι πληροφορίες για τις εναλλακτικές επιλογές super markets. Στο τέταρτο φύλλο, Answers, καταχωρούνται όλες οι απαντήσεις των ερωτώμενων για κάθε ένα κριτήριο και στις τελευταίες στήλες καταχωρείται η διάταξη των εναλλακτικών που έδωσε κάθε ερωτώμενος.

Εικόνα 4-16: Αρχείο εισόδου στο MARKEX

4.4.2. Εφαρμογή UTASTAR

Όταν το αρχείο εισαχθεί στο MARKEX, εφαρμόζεται η UTASTAR. Στη συνέχεια στο πεδίο “Optimization type” επιλέγεται το “Post optimization” προκειμένου να υπολογιστεί η βέλτιστη λύση των αποτελεσμάτων. Από την καρτέλα Consumer analysis και συγκεκριμένα στο Average Values εμφανίζονται όλα τα βάρη κριτηρίων καθώς και η μέση ολική χρησιμότητα όλων των εναλλακτικών.

Τα αποτελέσματα που δίνονται είναι οι ολικές χρησιμότητες για κάθε αποφασίζοντα και τα βάρη των κριτηρίων.

Παρακάτω παρουσιάζονται τα μέσα βάρη των κριτηρίων (βλ. πίνακα 4-2).

Πίνακας 4-2: Μέσα βάρη κριτηρίων

Τιμές προϊόντων	Προσφορές	Καθαριότητα και άνεση	Ποιότητα προϊόντων	Ποικιλία προϊόντων
0.076465	0.052128	0.051154	0.048811	0.047150

Απόσταση από το σπίτι	Δυνατότητα διανομής προϊόντων	Εξυπηρέτηση πελατών	Ασφάλεια στις συναλλαγές	Ύπαρξη εξειδικευμένων προϊόντων
0.060021	0.046175	0.045828	0.043152	0.046809

Ύπαρξη επαρκούς ποσότητας εμπορευμάτων	Ύπαρξη ποικιλίας βιολογικών προϊόντων	Ύπαρξη προϊόντων που μόλις έχουν κυκλοφορήσει	Ύπαρξη χώρου στάθμευσης	Ωράριο καταστήματος
0.044541	0.042923	0.044574	0.044750	0.042616

Ύπαρξη πολλών ταμείων σε λειτουργία	Διαφήμιση	Φήμη καταστήματος	Ύπαρξη τοπικών προϊόντων	Κάρτα bonus	Δυνατότητα αλλαγής ή επιδιόρθωσης ελαττωματικών προϊόντων
0.043350	0.044632	0.043853	0.043937	0.043395	0.043740

Το σημαντικότερο κριτήριο είναι οι τιμές των προϊόντων (0.076465) και δεύτερο σε βάρος κριτήριο είναι η απόσταση από το σπίτι (0.060021). Το κριτήριο με το μικρότερο μέσο βάρος είναι το ωράριο καταστήματος (0.042616).

Οι εναλλακτικές που χρησιμοποιήθηκαν ήταν τα super markets των εταιριών: AB Βασιλόπουλος, Σκλαβενίτης, Χαλκιαδάκης, LIDL, SYNKA.

Παρακάτω παρατίθεται ο πίνακας με την μέση ολική χρησιμότητα της κάθε εναλλακτικής (βλ. Πίνακα 4-3) για τους 199 ερωτώμενους βάσει των αποτελεσμάτων που εξήχθησαν από τη UTASTAR που εφαρμόστηκε στο λογισμικό MARKEKX.

Πίνακας 4-3: Μέση ολική χρησιμότητα εναλλακτικών

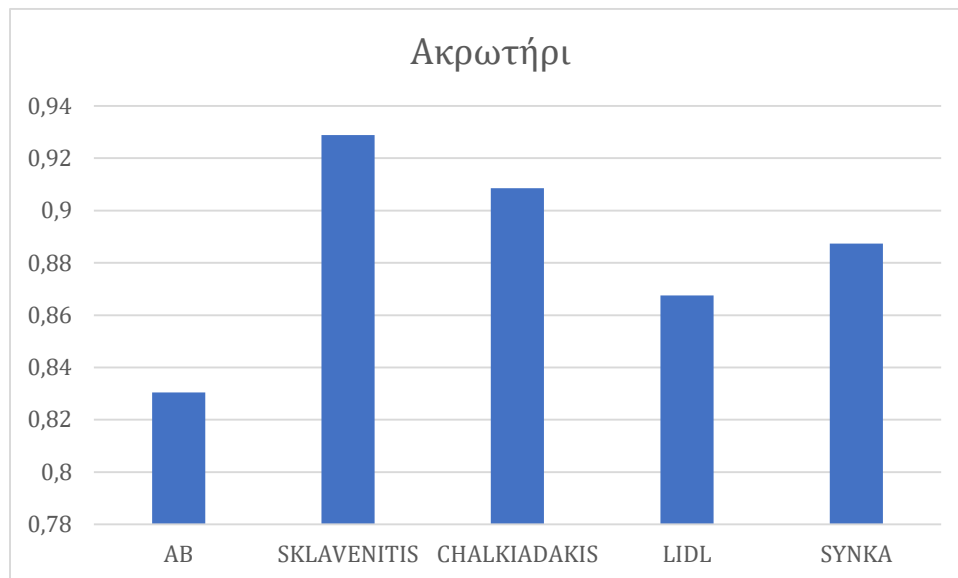
AB Βασιλόπουλος	Σκλαβενίτης	Χαλκιαδάκης	LIDL	SYNKA
0.844027	0.885857	0.839879	0.832342	0.861931

Την μεγαλύτερη μέση ολική χρησιμότητα έχει το super market της εταιρίας «Σκλαβενίτης», δεύτερο έρχεται το super market της εταιρίας «SYNKA», τρίτο αυτό της «AB Βασιλόπουλος», τέταρτο της «Χαλκιαδάκης» και τελευταίο αυτό της εταιρίας «LIDL».

4.4.2.1. Μέση ολική χρησιμότητα εναλλακτικών βάσει περιοχής

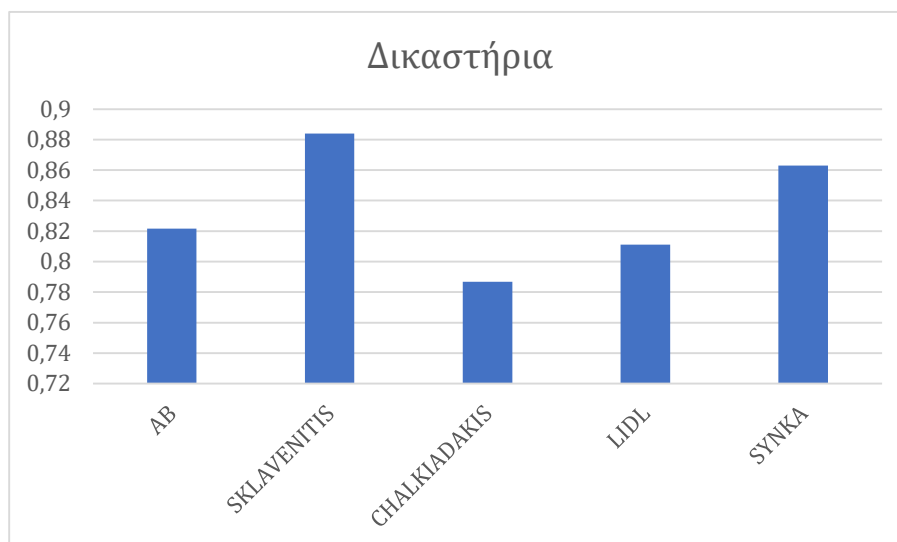
Υπολογίστηκε η μέση ολική χρησιμότητα των εναλλακτικών βάσει των περιοχών που δήλωσαν οι ερωτηθέντες ότι κατοικούν. Ο πίνακας των ολικών χρησιμοτήτων κάθε ερωτώμενου ταξινομήθηκε με βάση τις περιοχές που δηλώθηκαν. Για τις περιοχές με

τους περισσότερους ερωτώμενους, οι οποίες είναι: Ακρωτήρι, Δαράτσο, Δικαστήρια, Κέντρο, Λενταριανά, Νέα Χώρα, Σούδα και Χαλέπα, υπολογίστηκαν οι μέσοι όροι των ολικών χρησιμότητων των εναλλακτικών και παρουσιάζονται παρακάτω.



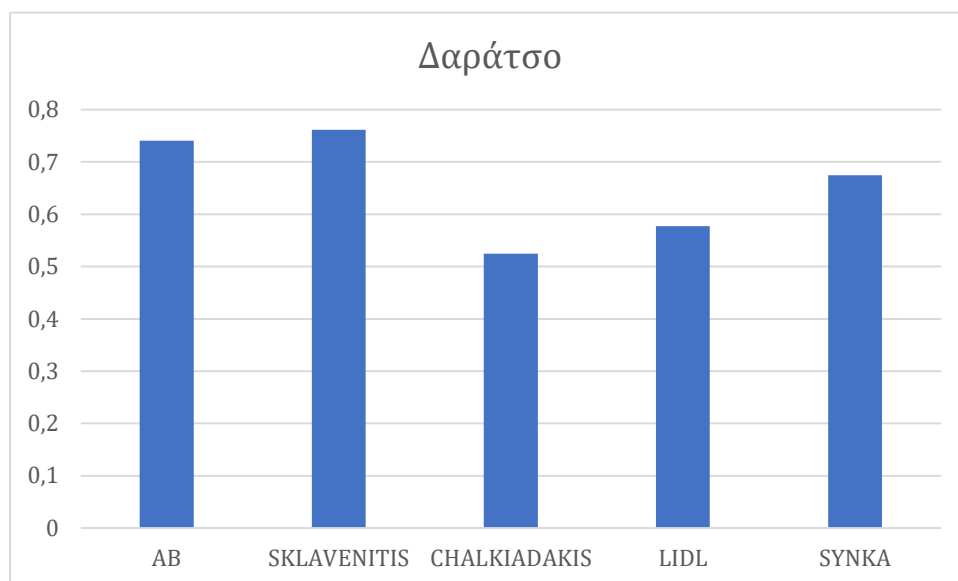
Γράφημα 4-15: Μέση ολική χρησιμότητα εναλλακτικών κατοίκων Ακρωτηρίου

Οι κάτοικοι της περιοχής του Ακρωτηρίου προτιμούν τα super markets «Σκλαβενίτης» και έπειτα τα super markets «Χαλκιαδάκης», ενώ τελευταίο σε προτίμηση είναι αυτό της εταιρίας «LIDL».



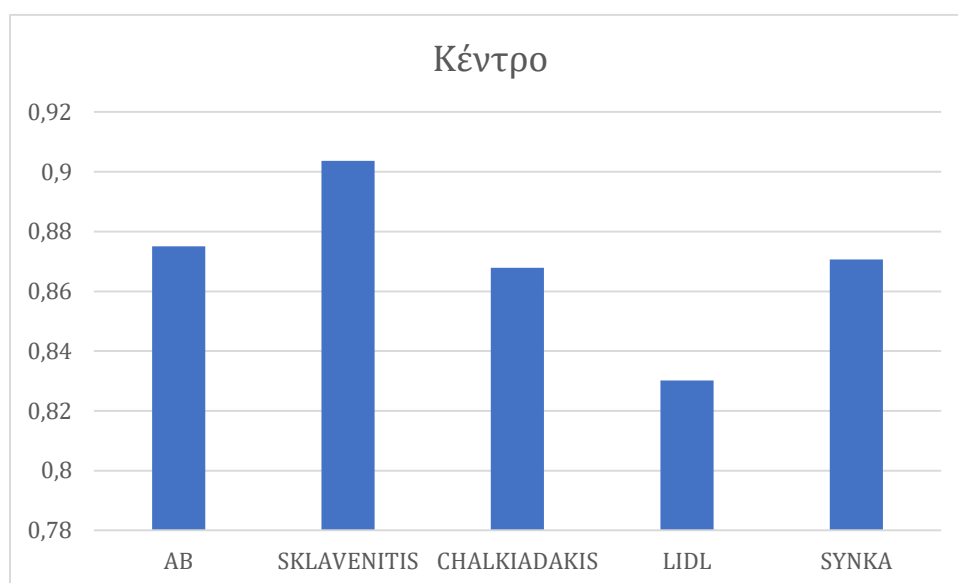
Γράφημα 4-16: Μέση ολική χρησιμότητα εναλλακτικών κατοίκων Δικαστηρίων

Οι κάτοικοι της περιοχής των Δικαστηρίων προτιμούν τα super markets «Σκλαβενίτης» και έπειτα τα super markets «SYNKA», ενώ τελευταίο σε προτίμηση είναι αυτό της εταιρίας «Χαλκιαδάκης».



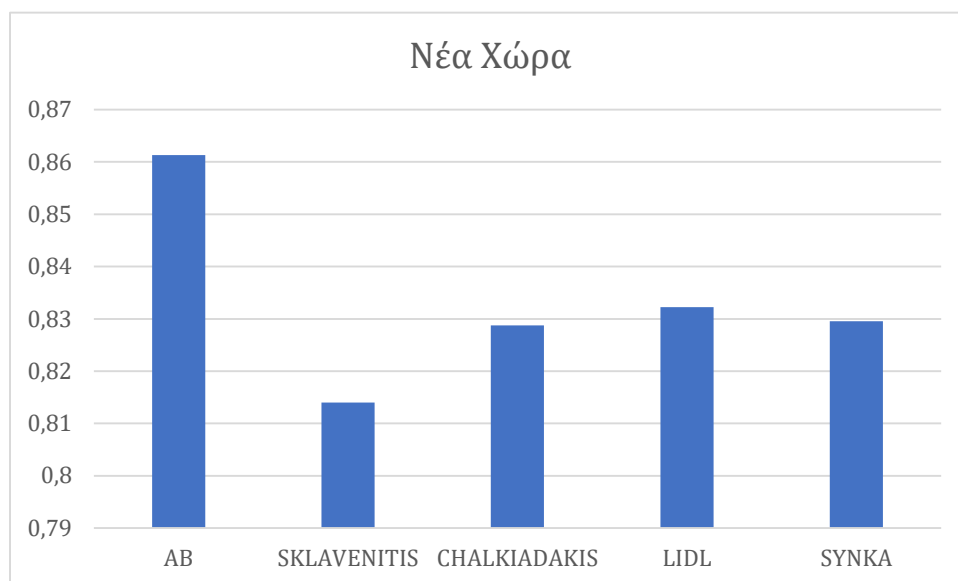
Γράφημα 4-17: Μέση ολική χρήςιμότητα εναλλακτικών κατοίκων Δαράτσου

Οι κάτοικοι της περιοχής του Δαράτσου προτιμούν τα super markets «Σκλαβενίτης» και έπειτα τα super markets «AB Βασιλόπουλος», ενώ τελευταίο σε προτίμηση είναι αυτό της εταιρίας «Χαλκιαδάκης».



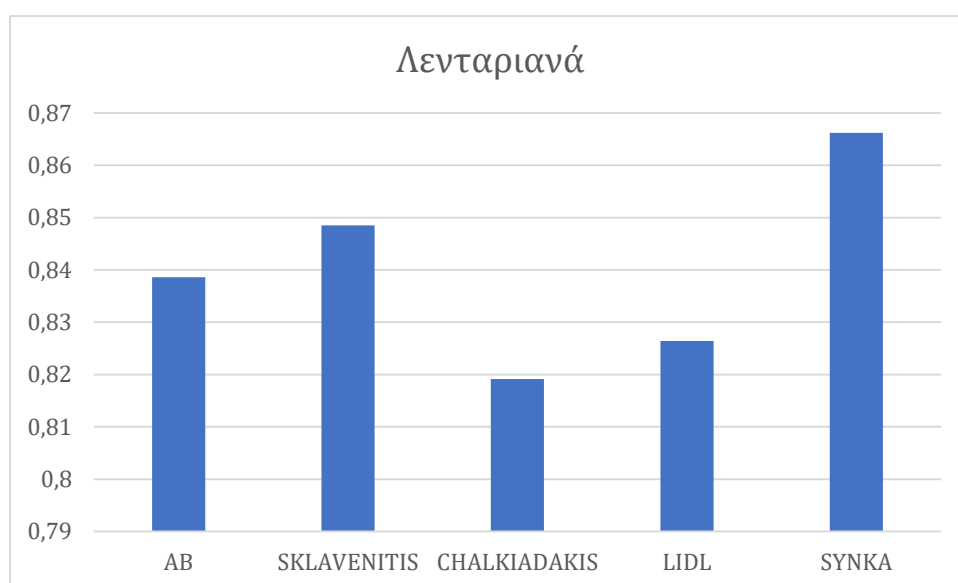
Γράφημα 4-18: Μέση ολική χρήςιμότητα εναλλακτικών κατοίκων Κέντρου

Οι κάτοικοι της περιοχής του Κέντρου προτιμούν τα super markets «Σκλαβενίτης» και έπειτα τα super markets «AB Βασιλόπουλος», ενώ τελευταίο σε προτίμηση είναι αυτό της εταιρίας «LIDL».



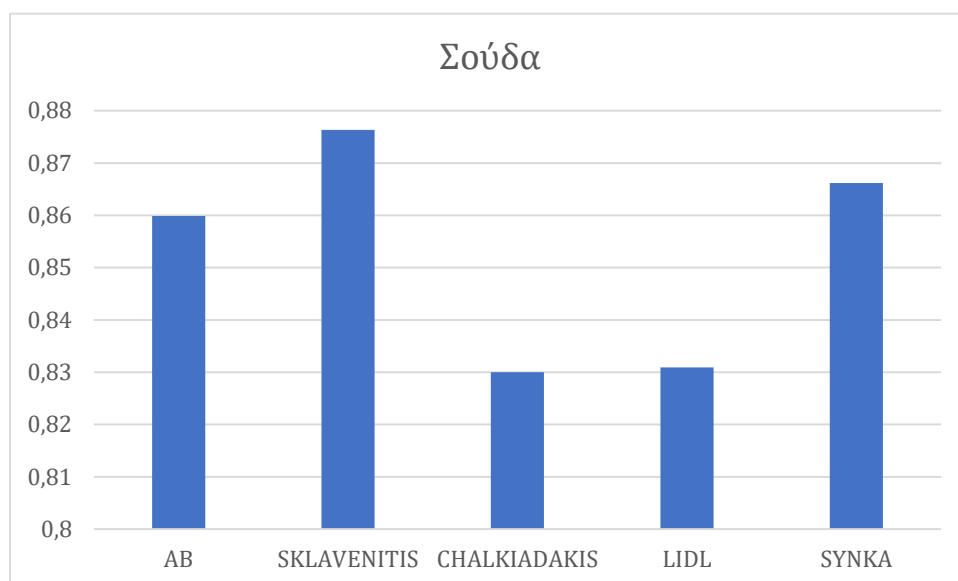
Γράφημα 4-19: Μέση ολική χρησιμότητα εναλλακτικών κατοίκων Νέας Χώρας

Οι κάτοικοι της περιοχής της Νέας Χώρας προτιμούν τα super markets «AB Βασιλόπουλος» και έπειτα τα super markets «LIDL», ενώ τελευταίο σε προτίμηση είναι αυτό της εταιρίας «Σκλαβενίτης».



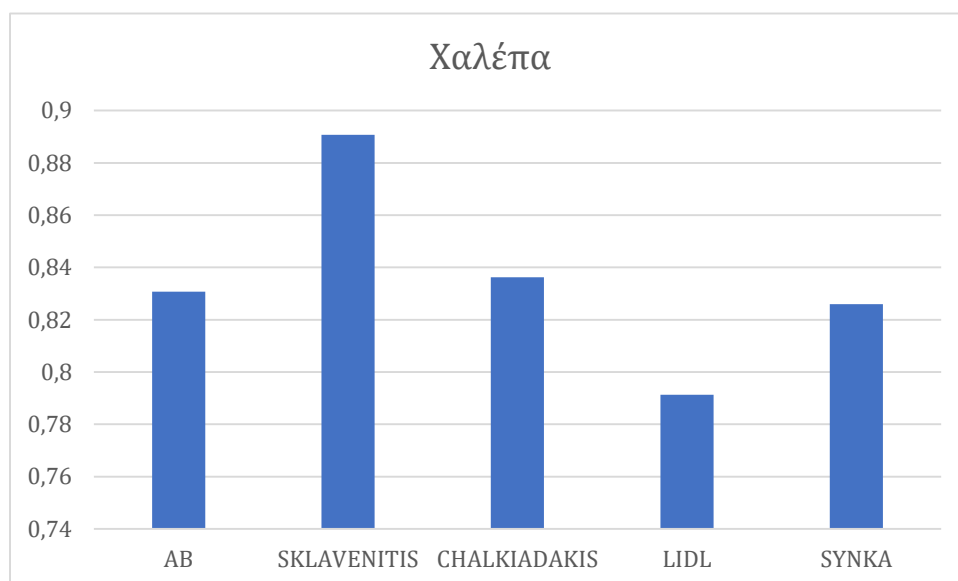
Γράφημα 4-20: Μέση ολική χρησιμότητα εναλλακτικών κατοίκων Λενταριανών

Οι κάτοικοι της περιοχής των Λενταριανών προτιμούν τα super markets «SYNKA» και έπειτα τα super markets «Σκλαβενίτης», ενώ τελευταίο σε προτίμηση είναι αυτό της εταιρίας «Χαλκιαδάκης».



Γράφημα 4-21: Μέση ολική χρησιμότητα εναλλακτικών κατοίκων Σούδας

Οι κάτοικοι της περιοχής της Σούδας προτιμούν τα super markets «Σκλαβενίτης» και έπειτα τα super markets «SYNKA», ενώ τελευταίο σε προτίμηση είναι αυτό της εταιρίας «Χαλκιαδάκης».



Γράφημα 4-22: Μέση ολική χρησιμότητα εναλλακτικών κατοίκων Χαλέπας

Οι κάτοικοι της περιοχής της Χαλέπας προτιμούν τα super markets «Σκλαβενίτης» και έπειτα τα super markets «Χαλκιαδάκης», ενώ τελευταίο σε προτίμηση είναι αυτό της εταιρίας «LIDL».

Συμπεράσματα UTASTAR

Από τα αποτελέσματα της πολυκριτήριας ανάλυσης UTASTAR προκύπτει ότι από τα μέσα βάρη κριτηρίων, το σημαντικότερο κριτήριο είναι αυτό της τιμής των προϊόντων,

ενώ δεύτερο σημαντικότερο είναι το κριτήριο της απόστασης του super market από το σπίτι. Όσον αφορά στην μέση ολική χρησιμότητα των εναλλακτικών, το super market με την μεγαλύτερη μέση ολική χρησιμότητα είναι αυτό της εταιρίας «Σκλαβενίτης». Στη συνέχεια πραγματοποιήθηκε η ανάλυση της μέσης ολικής χρησιμότητας βάσει των περιοχών που δήλωσαν οι ερωτώμενοι ότι διαμένουν. Η ανάλυση αυτή πραγματοποιήθηκε για τις περιοχές Ακρωτήρι, Δαράτσο, Δικαστήρια, Κέντρο, Λενταριανά, Νέα Χώρα, Σούδα και Χαλέπα και όχι για το σύνολο των περιοχών, καθώς αυτές ήταν οι περιοχές με τις περισσότερες απαντήσεις. Επομένως, οι κάτοικοι των περιοχών Ακρωτηρίου, Δαράτσου, Δικαστηρίων, Κέντρου, Σούδας και Χαλέπας προτιμούν το super market «Σκλαβενίτης», ενώ οι κάτοικοι της περιοχής της Νέας Χώρας προτιμούν το super market «ΑΒ Βασιλόπουλος» και τέλος, οι κάτοικοι της περιοχής των Λενταριανών προτιμούν το super market «SYNKA».




4.5. ΣΥΣΤΑΔΟΠΟΙΗΣΗ

4.5.1. Συσταδοποίηση με χρήση αλγορίθμων MARKECH

Προκειμένου να δημιουργηθούν οι κατάλληλες συστάδες μέσω της συσταδοποίησης στο MARKECH χρησιμοποιήθηκαν τα κριτήρια που κρίθηκαν πιο σημαντικά από τους καταναλωτές. Ο αριθμός των καταναλωτών που προκύπτει σε κάθε συστάδα έχει κατώφλι κριτηρίου μεγαλύτερο από αυτό που έχει θεωρηθεί σημαντικό στην συγκεκριμένη συστάδα και αντίστοιχα μικρότερο κατώφλι κριτηρίου για τα κριτήρια που δεν θεωρούνται σημαντικά. Στόχος της συγκεκριμένης συσταδοποίησης είναι η δημιουργία συστάδων οι οποίες θα καλύπτουν τα σημαντικά κριτήρια του μεγαλύτερου δυνατού εύρους καταναλωτών. Στον παρακάτω πίνακα (βλ. πίνακα 4-4) παρουσιάζονται τα αποτελέσματα της συσταδοποίησης, δηλαδή το ποσοστό των καταναλωτών που συμπεριρίζεται την σημαντικότητα ή όχι των αντίστοιχων κριτηρίων, ανάλογα με το κατώφλι που δίνεται για κάθε κριτήριο. Επί παραδείγματι, στο Cluster 0 συμπεριλαμβάνονται οι καταναλωτές που τους αναλογεί τιμή πάνω από 0,076 για το κριτήριο της τιμής των προϊόντων και ταυτόχρονα τιμή μικρότερη από 0,047 για το κριτήριο της ύπαρξης εξειδικευμένων προϊόντων, εφόσον αυτό δεν θεωρείται σημαντικό.

Πίνακας 4-4: Συστάδες καταναλωτών με βάση τα κριτήρια επιλογής super market

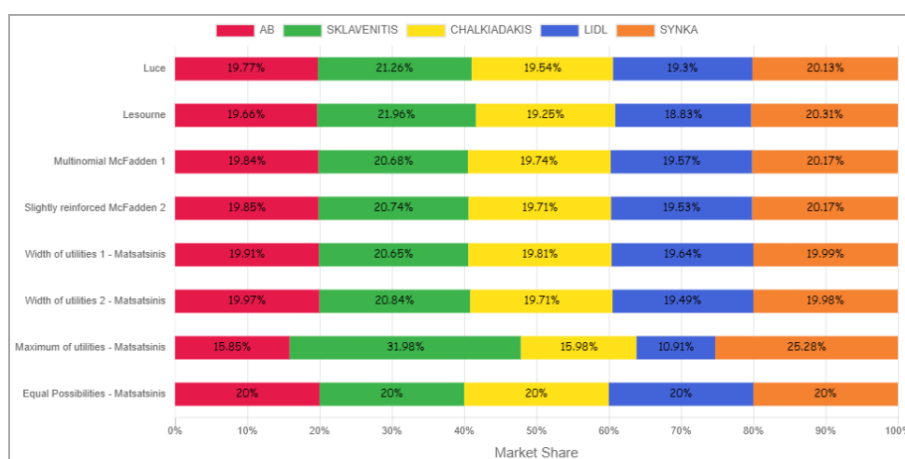
Created Clusters

Cluster Name	Criteria	Analysis Method	Consumers Number	Actions
Cluster 0	Products_Prices (0.076, Significant)	Average Weights	66 (33.17%)	
	Specialized_Products (0.047, Non Significant)			
Cluster 1	Products_Quality (0.049, Non Significant)	Average Weights	75 (37.69%)	
	Quantity (0.045, Significant)			
Cluster 2	Offers (0.052, Non Significant)	Average Weights	65 (32.66%)	
	Products_Quality (0.049, Non Significant)			
	Bonus_card (0.043, Significant)			

Επομένως, γίνεται αντιληπτό ότι μία αύξηση των κριτηρίων που κρίνονται σημαντικά και είναι δυνατό να επιφέρει μία αντίστοιχη αύξηση ικανοποίησης των καταναλωτών και πιθανότατα αύξηση του καταναλωτικού κοινού. Συγκεκριμένα, το 33,71% των ερωτώμενων θεωρούν ότι η τιμή των προϊόντων είναι ένα σημαντικό κριτήριο και ταυτόχρονα ότι η ύπαρξη εξειδικευμένων προϊόντων δεν είναι σημαντική. Επίσης, ένα μεγαλύτερο ποσοστό της τάξης του 37,69% θεωρεί ότι είναι σημαντική η επαρκής ποσότητα των προϊόντων, αλλά όχι η ποιότητά τους. Τέλος, ένα ποσοστό 32,66% των ερωτώμενων θεωρεί ότι η ύπαρξη προσφορών και κάρτας bonus είναι σημαντικά κριτήρια, ενώ δεν θεωρεί σημαντικό κριτήριο την ποιότητα των προϊόντων.

Με βάση τα παραπάνω, τα μερίδια αγοράς σύμφωνα με την προσομοίωση της αγοράς από το λογισμικό MARKEKX έχουν ως εξής.

Τα αρχικά μερίδια αγοράς των super market του Ν. Χανίων παρουσιάζονται παρακάτω (Γράφημα 4-23)



Γράφημα 4-23: Αρχικά μερίδια αγοράς

Όπως παρατηρείται, το μεγαλύτερο ποσοστό στην αγορά έχει το super market «Σκλαβενίτης» και το μικρότερο το super market «LIDL».

Στην συνέχεια παρουσιάζονται τα αποτελέσματα των μεριδίων της αγοράς με την τροποποίηση των αντίστοιχων κριτηρίων που έδωσε η συσταδοποίηση των κριτηρίων από το MARKEKX.

Αρχικά η πρώτη συστάδα, η οποία αναφέρει ότι οι καταναλωτές επιλέγουν ως σημαντικό κριτήριο τις τιμές των προϊόντων, ενώ ως ασήμαντο την ύπαρξη εξειδικευμένων προϊόντων έχει ως αποτέλεσμα το παρακάτω γράφημα (Γράφημα 4-24)



Γράφημα 4-24: Μεριδία αγοράς 1^{ης} συστάδας

Το super market «Σκλαβενίτης» συνεχίζει να έχει το μεγαλύτερο μερίδιο αγοράς και πιο συγκεκριμένα με τα κριτήρια που έχει θέσει ως σημαντικά και ασήμαντα η συγκεκριμένη συστάδα καταναλωτών, φαίνεται να υπάρχει μία αύξηση του μεριδίου της αγοράς. Αύξηση του μεριδίου της αγοράς παρουσιάζει και το super market «Χαλκιαδάκης» στην συγκεκριμένη συστάδα καταναλωτών.

Για την 2^η συσταδα τα αποτελέσματα παρατίθενται στο παρακάτω γράφημα (Γράφημα 4-25)

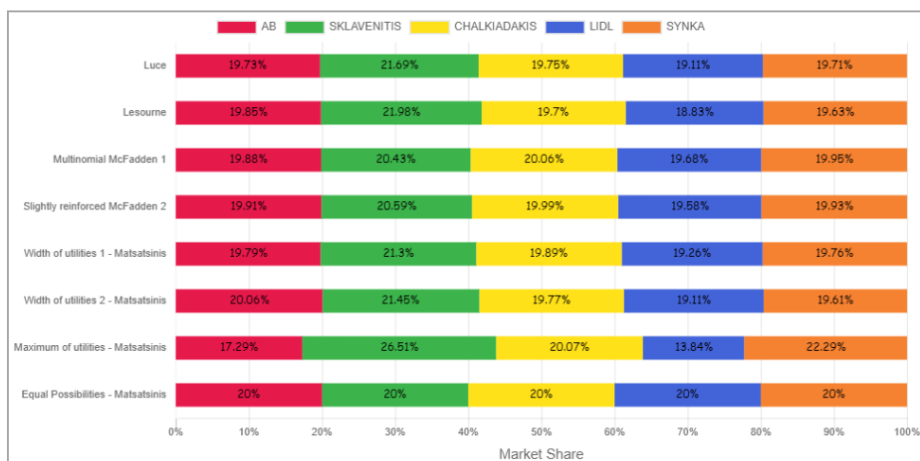
All Models

Cluster 1

Present Market

New Scenario

Graphical



Γράφημα 4-25: Μερίδια αγοράς 2^{ης} συστάδας

Όπως παρατηρείται, ξανά το super market με το μεγαλύτερο μερίδιο αγοράς είναι το «Σκλαβενίτης» και ακολουθούν «Χαλκιαδάκης», «AB Βασιλόπουλος», «LIDL» και «SYNKA».

Τέλος, τα αποτελέσματα της 3^{ης} συστάδας καταναλωτών, όπως προέκυψαν από την προσομοίωση αγοράς φαίνονται στο παρακάτω γράφημα. (Γράφημα 4-26)

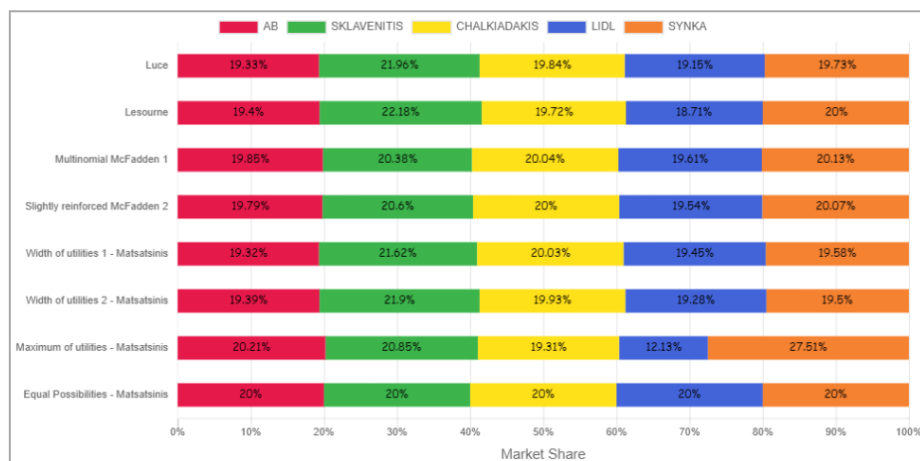
All Models

Cluster 2

Present Market

New Scenario

Graphical



Γράφημα 4-26: Μερίδια αγοράς 3^{ης} συστάδας

Στην συγκεκριμένη συστάδα, ενώ το super market «Σκλαβενίτης» είναι ξανά πρώτο, τα υπόλοιπα μερίδια διαφοροποιούνται και στην δεύτερη θέση βρίσκεται το super market «Χαλκιαδάκης», στην τρίτη το «SYNKA» και έπειτα το «AB Βασιλόπουλος» και τέλος, το «LIDL».

4.5.2. Αλγόριθμος Συσταδοποίησης K-means

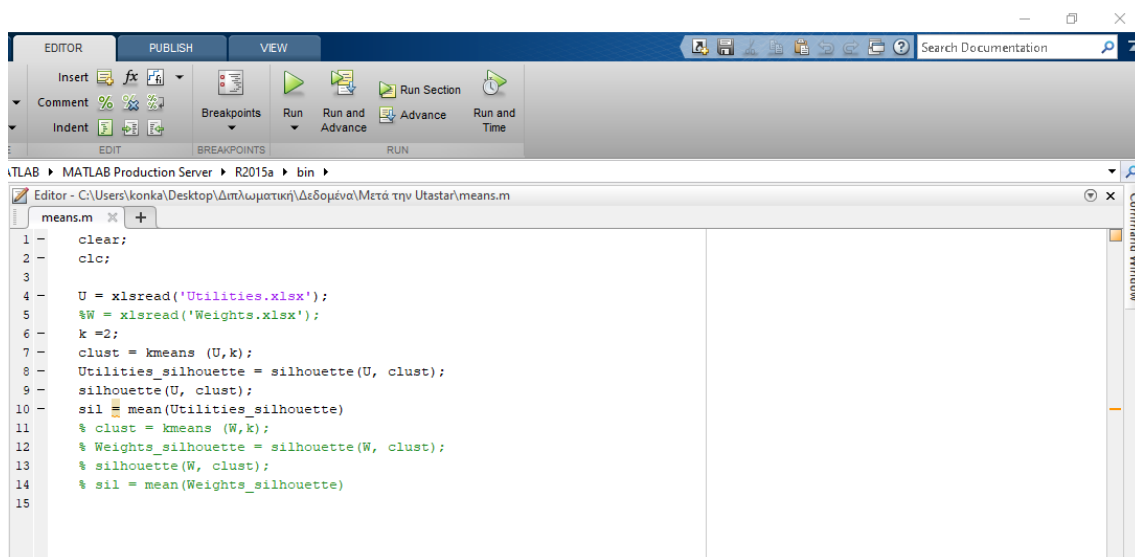
4.5.2.1. Εισαγωγή

Μετά την περιγραφική στατιστική, την προεπεξεργασία δεδομένων και την πολυκριτήρια ανάλυση UTASTAR θα ακολουθήσει η διαδικασία της συσταδοποίησης με χρήση του αλγορίθμου k-means. Στόχος της συγκεκριμένης διαδικασίας είναι να ομαδοποιηθούν οι καταναλωτές και να εξεταστεί το προφίλ τους βάσει των κριτηρίων και των εναλλακτικών που επιλέγουν. Τα δεδομένα που πρόκειται να χρησιμοποιηθούν για την συσταδοποίηση προέρχονται από τα αποτελέσματα της πολυκριτήριας ανάλυσης. Συγκεκριμένα, χρησιμοποιούνται τα δεδομένα από την κατηγορία Global Utilities καθώς και τα δεδομένα από την κατηγορία Criteria Weights, όπως υπολογίστηκαν στο MARKEX. Ο πίνακας από την κατηγορία Global Utilities που χρησιμοποιήθηκε αποτελούνταν από 5 εναλλακτικές επιλογές (τα αντίστοιχα super markets) και 199 αναφορές, όσοι δηλαδή οι ερωτώμενοι. Τα αποτελέσματα αντιστοιχούσαν στις προτιμήσεις των εναλλακτικών για κάθε ερωτώμενο. Αντίστοιχα, ο πίνακας από την κατηγορία Criteria Weights αποτελούνταν από 21 κριτήρια και 199 αναφορές. Τα αποτελέσματα αντιστοιχούσαν στο βάρος του εκάστοτε κριτηρίου για κάθε ερωτώμενο.

4.5.2.2. Εφαρμογή του αλγορίθμου k-means

Για την εφαρμογή του αλγορίθμου k-means χρησιμοποιήθηκε το λογισμικό MATLAB. Χρησιμοποιήθηκαν δύο διαφορετικά αρχεία, ένα αρχείο "Utilities" και ένα αρχείο "Weights". Αρχικά, και για τα δύο αρχεία χωριστά, ελέγχθηκε ο δείκτης Silhouette προκειμένου να ελεγχθεί κατά πόσο οι συστάδες περιέχουν στοιχεία που μοιάζουν μεταξύ τους και που ταυτόχρονα είναι ανόμοια με τα στοιχεία των υπόλοιπων συστάδων. Δοκιμάστηκε διαφορετικός αριθμός συστάδων και για τα δύο αρχεία με σκοπό την εύρεση του βέλτιστου αριθμού συστάδων. Ο αρχικός αριθμός συστάδων που δόθηκε ήταν $k=2$. Στην συνέχεια εφαρμόστηκε η συνάρτηση k-means προκειμένου να βρεθούν τα στοιχεία που ανήκουν σε κάθε συστάδα. Για κάθε αρχείο χωριστά, δόθηκε ο αριθμός συστάδων που βρέθηκε ως βέλτιστος στο προηγούμενο βήμα.

Αρχικά το αρχείο που χρησιμοποιήθηκε για τον δείκτη Silhouette (βλ. παρ. 3.6) φαίνεται στην εικόνα 4-16.



Εικόνα 4-17: Αρχείο Matlab για εύρεση δείκτη Silhouette

Τα αποτελέσματα που δόθηκαν για τα αρχεία “Utilities” και “Weights” εμφανίζονται στους παρακάτω πίνακες (πίνακας 4-5 και 4-6).

Πίνακας 4-5: Δείκτης Silhouette για το αρχείο “Utilities”

Αριθμός συστάδων	Δείκτης Silhouette
2	0,9925
3	1
4	0,7236

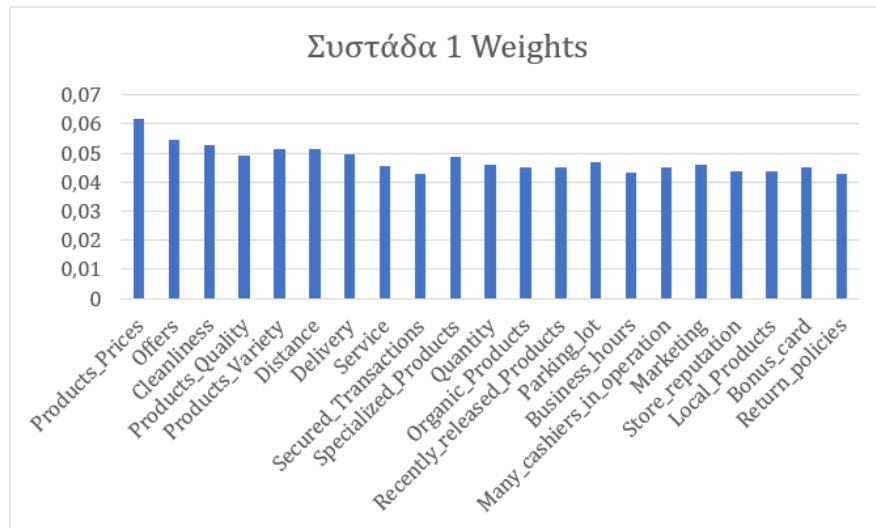
Πίνακας 4-6: Δείκτης Silhouette για το αρχείο “Weights”

Αριθμός συστάδων	Δείκτης Silhouette
2	0,3599
3	0,4959
4	0,2609
5	0,2479
6	0,4265
7	0,2141

Στο αρχείο “Utilities” η καλύτερη συσταδοποίηση επιτυγχάνεται με αριθμό συστάδων $k=3$ και με δείκτη Silhouette 1, ο οποίος επιβεβαιώνει μία καλή συσταδοποίηση. Στο αρχείο “Weights” η καλύτερη συσταδοποίηση επιτυγχάνεται με αριθμό συστάδων $k=3$ και με δείκτη Silhouette 0,4959.

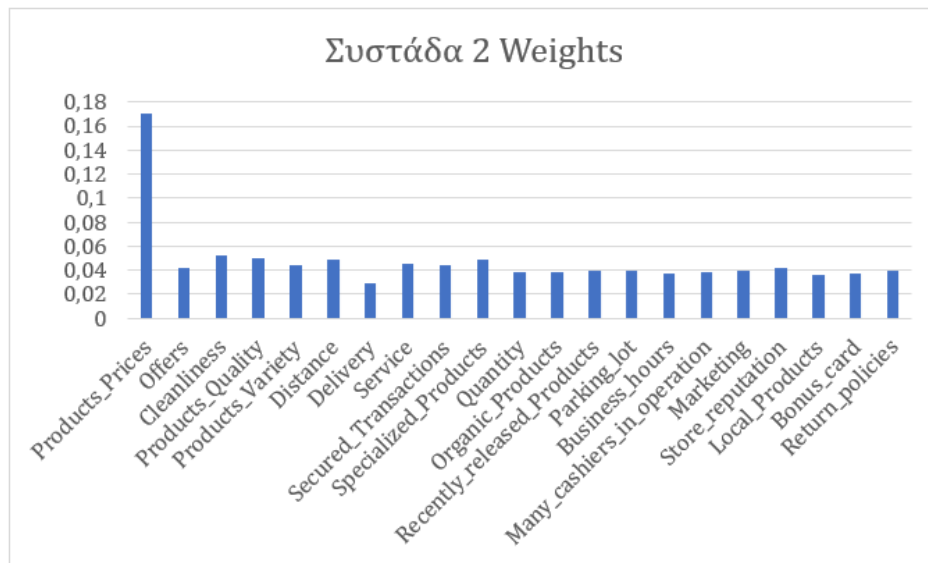
4.5.2.3. Αποτελέσματα αλγορίθμου *k-means* για τα βάρη των κριτηρίων

Τα αποτελέσματα που θα αναλυθούν παρακάτω προέρχονται από την συσταδοποίηση που έδωσε ο αλγόριθμος *k-means* μέσω της γλώσσας προγραμματισμού Matlab. Στη συνέχεια διαχωρίστηκαν οι καταναλωτές που ανήκουν σε κάθε συστάδα και κατόπιν βρέθηκε ο μέσος όρος των βαρών των κριτηρίων και δημιουργήθηκαν τα παρακάτω γραφήματα (Γράφημα 4-27, 4-28 και 4-29) για τις 3 συστάδες.



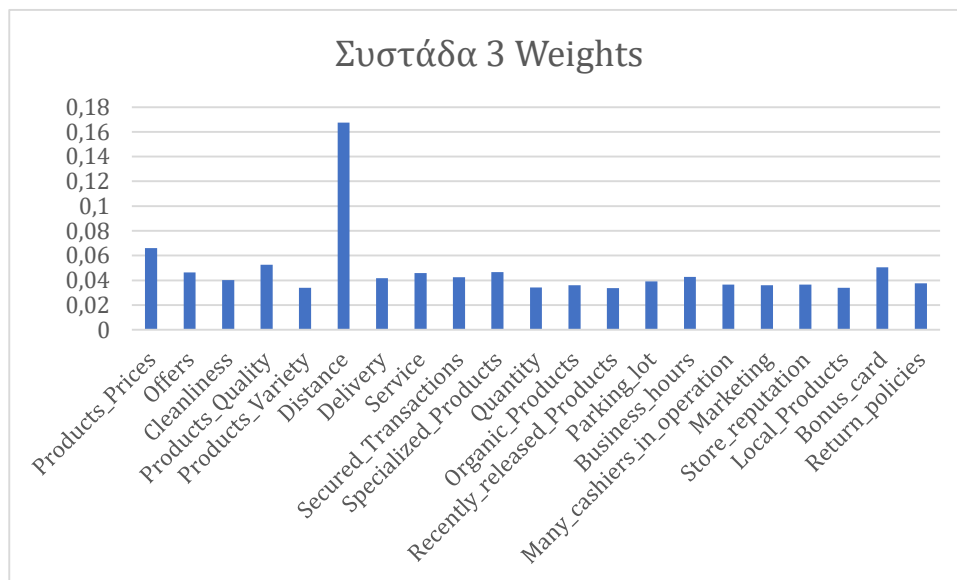
Γράφημα 4-27: Βάρη κριτηρίων- Συστάδα 1

Στην πρώτη συστάδα το κριτήριο με το μεγαλύτερο βάρος είναι αυτό της τιμής των προϊόντων και ακολουθεί αυτό των προσφορών, καθώς και αυτό της καθαριότητας και άνεσης του super market. Κοντά σε αυτά τα κριτήρια είναι η ποικιλία και η ποιότητα των προϊόντων. Μικρές είναι οι διαφορές που παρατηρούνται ανάμεσα στα περισσότερα κριτήρια, με τα κριτήρια που έχουν τα μικρότερα βάρη να είναι η ασφάλεια στις συναλλαγές, το ωράριο του καταστήματος και οι πολιτικές επιστροφών.



Γράφημα 4-28: Βάρη κριτηρίων-Συστάδα 2

Στο παραπάνω γράφημα (Γράφημα 4-28) που αφορά στη συστάδα 2 για τα βάρη των κριτηρίων, παρατηρείται ότι σχεδόν όλα τα κριτήρια βρίσκονται στο ίδιο επίπεδο εκτός από το κριτήριο με την μεγαλύτερη βαρύτητα που είναι αυτό της τιμής των προϊόντων, το οποίο φαίνεται να έχει μεγάλη διαφορά στην τιμή του από τα υπόλοιπα κριτήρια. Αντίστοιχα, το κριτήριο με την μικρότερη βαρύτητα είναι αυτό της διανομής προϊόντων στο σπίτι.

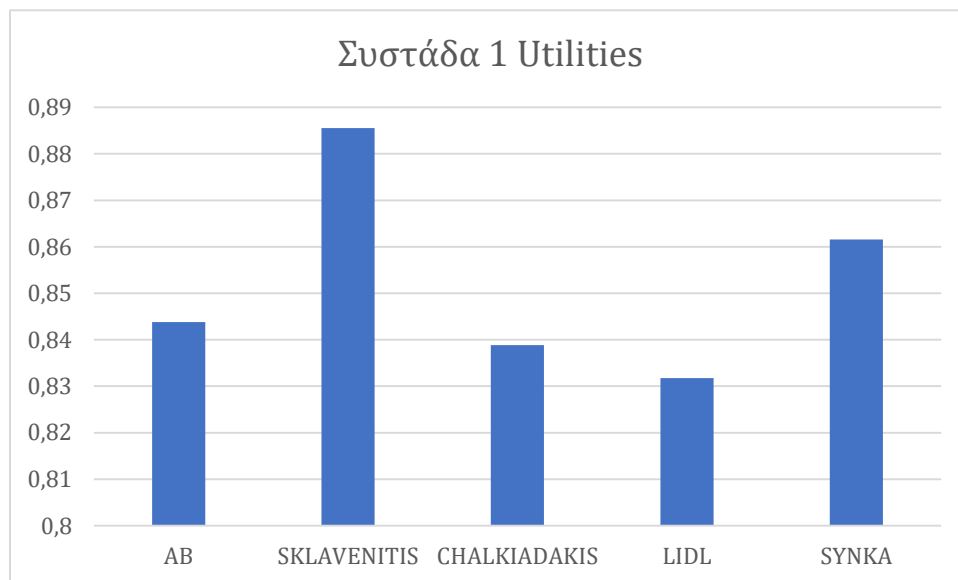


Γράφημα 4-29: Βάρη κριτηρίων-Συστάδα 3

Στο γράφημα 4-29 γίνεται ξεκάθαρο ότι το κριτήριο των καταναλωτών που ανήκουν στην συστάδα 3 με το μεγαλύτερο βάρος είναι αυτό της απόστασης από το super market. Επόμενο στη σειρά έρχεται αυτό της τιμής των προϊόντων και της ποιότητας των προϊόντων. Τα υπόλοιπα κριτήρια βρίσκονται σε πολύ κοντινές τιμές μεταξύ τους.

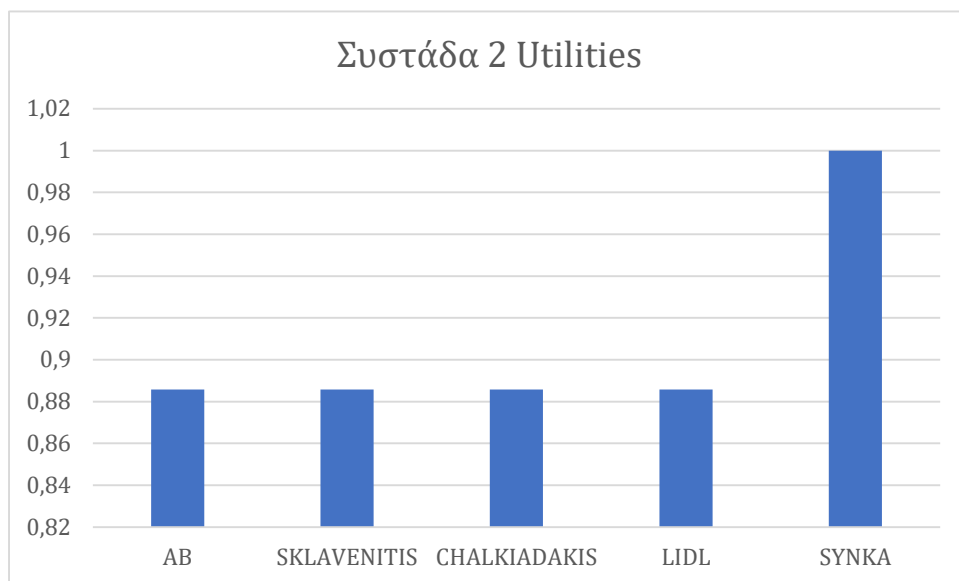
4.5.2.4. Αποτελέσματα αλγορίθμου *k-means* για τις ολικές χρησιμότητες των εναλλακτικών επιλογών

Τα αποτελέσματα που θα αναλυθούν παρακάτω αντλήθηκαν από τις συστάδες που δημιούργησε ο αλγόριθμος *k-means* μέσω της γλώσσας προγραμματισμού Matlab. Αφού διαχωρίστηκαν, βρέθηκαν οι μέσοι όροι των ολικών χρησιμοτήτων των εναλλακτικών επιλογών κάθε συστάδας.



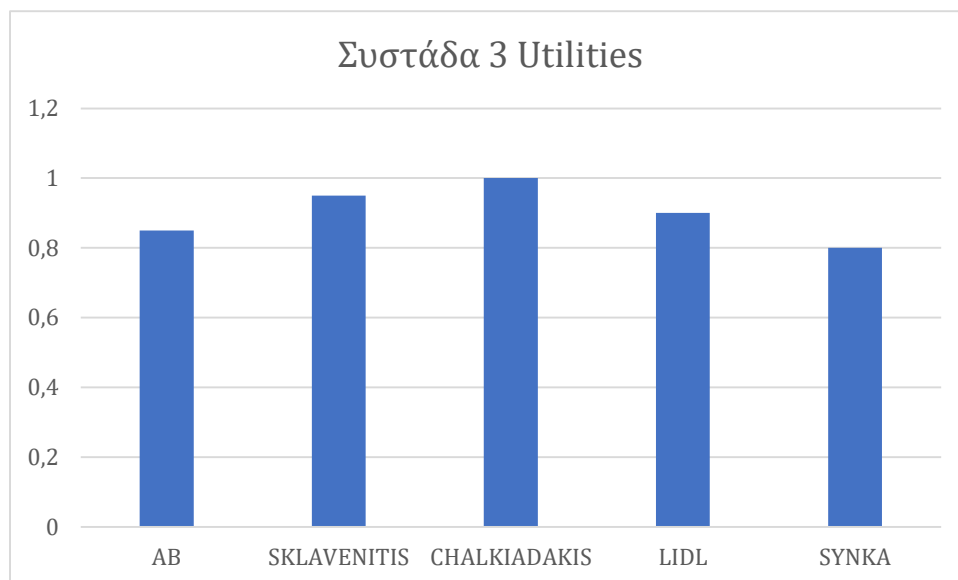
Γράφημα 4-30: Ολικές χρησιμότητες εναλλακτικών επιλογών-Συστάδα 1

Για την πρώτη συστάδα που δημιουργήθηκε, οι καταναλωτές επιλέγουν ως πρώτο σε προτίμηση super market αυτό της εταιρίας «Σκλαβενίτης» με ολική χρησιμότητα 0,8855 και ακολουθούν «SYNKA» με ολική χρησιμότητα 0,8615, «AB Βασιλόπουλος» με 0,8438, «Χαλκιαδάκης» με 0,8388 και τελευταίο αυτό της «LIDL» με 0,8317. Συμπεραίνεται, ότι οι καταναλωτές τις πρώτης συστάδας πρώτα θα επιλέξουν να ψωνίσουν από το Σκλαβενίτη και στη συνέχεια από το SYNKA και τον AB Βασιλόπουλο.



Γράφημα 4-31: Ολικές χρησιμότητες εναλλακτικών επιλογών-Συστάδα 2

Οι ολικές χρησιμότητες που υπολογίστηκαν για την συστάδα 2 δείχνουν ξανά το super market «SYNKA» στην πρώτη θέση με ολική χρησιμότητα 1, ενώ όλα τα υπόλοιπα super market βρίσκονται στην δεύτερη θέση με ολική χρησιμότητα 0,8858.



Γράφημα 4-32: Ολικές χρησιμότητες εναλλακτικών επιλογών-Συστάδα 3

Για την τρίτη συστάδα που δημιουργήθηκε, οι καταναλωτές επιλέγουν ως πρώτο σε προτίμηση super market αυτό της εταιρίας «Χαλκιαδάκης» με ολική χρησιμότητα 1 και ακολουθούν «Σκλαβενίτης» με ολική χρησιμότητα 0,95, «LIDL» με 0,90, «AB Βασιλόπουλος» με 0,85 και τελευταίο αυτό της «SYNKA» με 0,80. Συμπεραίνεται, ότι οι καταναλωτές τις πρώτης συστάδας πρώτα θα επιλέξουν να ψωνίσουν από το Χαλκιαδάκη και στη συνέχεια από το Σκλαβενίτη και το super market της εταιρίας «LIDL».

Συμπεράσματα Συσταδοποίησης

Όσον αφορά στη συσταδοποίηση, πραγματοποιήθηκαν δύο διαφορετικές συσταδοποιήσεις, μία με τη χρήση του MARKEX και μία με τη χρήση του αλγορίθμου k-means. Στην συσταδοποίηση μέσω του MARKEX δημιουργήθηκαν τρεις συστάδες με τα μεγαλύτερα δυνατά ποσοστά. Η πρώτη που αφορούσε το 33,71% των ερωτηθέντων θεωρεί σημαντικό το κριτήριο της τιμής, η δεύτερη (37,69%) το κριτήριο της ποσότητας και η τρίτη (32,66%) θεωρεί σημαντική την ύπαρξη κάρτας bonus. Για τα κριτήρια που δεν είναι σημαντικά, η πρώτη συστάδα καταναλωτών θεωρεί ότι η ύπαρξη εξειδικευμένων προϊόντων δεν είναι σημαντική, η δεύτερη θεωρεί ότι η ποιότητα των προϊόντων δεν είναι σημαντική, ενώ η τρίτη θεωρεί ασήμαντη την ύπαρξη προσφορών και ποιότητας προϊόντων. Στην συσταδοποίηση με χρήση αλγορίθμου k-means, δημιουργήθηκαν τρεις συστάδες καταναλωτών για τα βάρη των κριτηρίων και τρεις για τις ολικές χρησιμότητες των εναλλακτικών επιλογών. Για τα βάρη των κριτηρίων η πρώτη και η δεύτερη συστάδα επέλεξαν ως πιο σημαντικό κριτήριο τις τιμές των προϊόντων με μόνη διαφορά την κατάταξη των υπολοίπων κριτηρίων. Η τρίτη συστάδα θεωρεί σημαντικότερο το κριτήριο της απόστασης από το super market. Όσον αφορά στις ολικές χρησιμότητες, η πρώτη συστάδα προτείνει το super market «Σκλαβενίτης», η δεύτερη το «SYNKA» και η τρίτη το «Χαλκιαδάκης».

4.6. MARKET BASKET ANALYSIS

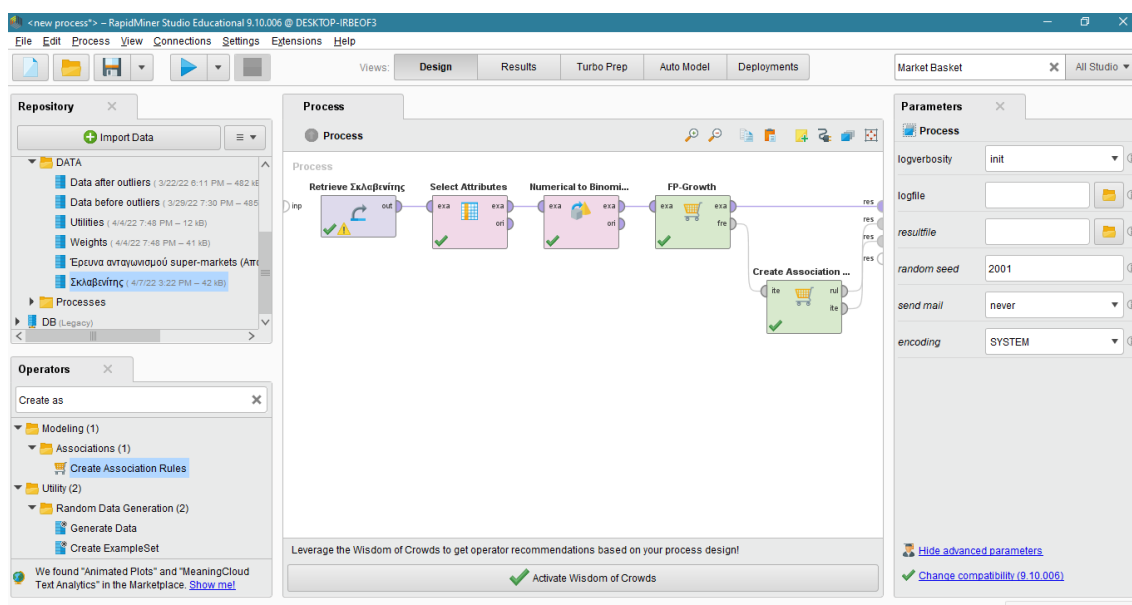
4.6.1. Εισαγωγή

Αρχικά, τα δεδομένα τα οποία αντλήθηκαν προκύπτουν από τις απαντήσεις της ερώτησης 5 του ερωτηματολογίου (παράρτημα). Ωστόσο, προκειμένου να εισαχθούν και να προκύψουν κανόνες συσχέτισης θα πρέπει πρώτα να μορφοποιηθούν κατάλληλα. Επομένως, για κάθε super market, καθώς και για την απάντηση «κανένα» που έχει δοθεί, πρέπει να δημιουργηθούν διαφορετικά αρχεία Excel, στα οποία θα προστεθούν όλες οι απαντήσεις όλων των καταναλωτών. Σε κάθε λογιστικό φύλλο Excel θα δοθεί η τιμή 1 για το super market που αντιστοιχεί στο συγκεκριμένο λογιστικό φύλλο, και η τιμή 0 για όλα τα υπόλοιπα.

Στο RapidMiner εισάγονται αρχικά τα δεδομένα με το τρόπο που έχει αναφερθεί παραπάνω (Παράγραφος 4.2.1) και στη συνέχεια δημιουργείται μία νέα διαδικασία. Τα δεδομένα εισάγονται στο πεδίο του “Design” και, όπως φαίνεται στην εικόνα 4-18, στη συνέχεια εισάγεται ο operator “Select Attributes”. Στον συγκεκριμένο operator στην καρτέλα Parameters επιλέγεται στο πεδίο attribute filter type το subset, προκειμένου να εισαχθεί μία ομάδα χαρακτηριστικών και όχι μόνο ένα. Έπειτα, προστίθεται ο operator “Numerical to Binomial” ώστε να μετατραπούν οι τιμές 0 και 1 του αρχείου

Excel σε τιμές True και False. Τέλος, εισάγονται οι operators “FP Growth” και “Create Association Rules” για την εκτέλεση του Market Basket Analysis. Ο operator “FP Growth” χρησιμοποιείται για να μετρήσει την συχνότητα που εμφανίζονται, και κατα συνέπεια αγοράζονται, τα διάφορα σύνολα προϊόντων από τους καταναλωτές. Γι’ αυτό στη συνέχεια συνδέεται και στον operator “Create Association Rules”, αλλά ταυτόχρονα και στα results. Στις παραμέτρους του “FP Growth” υπάρχει η δυνατότητα επιλογής του δείκτη support (υποστήριξη), ενώ στις παραμέτρους του “Create Association Rules” δίνεται η δυνατότητα για επιλογή κριτηρίου. Στην επιλογή του κριτηρίου confidence υπάρχει η δυνατότητα εισαγωγής ελάχιστης τιμής, ενώ για τον δείκτη lift η τιμή είναι 0.8 δοσμένη από το σύστημα.

Επιλέχθηκε να εφαρμοστεί αλγόριθμος «FP Growth» διότι, κατά την αναζήτηση του αλγορίθμου για την εκτέλεση του MBA στο λογισμικό RapidMiner, διαπιστώθηκε πως δεν υπήρχε άλλος διαθέσιμος αλγόριθμος για την εξαγωγή των κανόνων συσχέτισης.



Εικόνα 4-16: Σχεδιασμός της διαδικασίας εκτέλεσης για το Market Basket Analysis

4.6.2. Αποτελέσματα Market Basket Analysis για το Super Market «AB Βασιλόπουλος»

Αφού ολοκληρώθηκαν όλες οι παραπάνω διεργασίες, δημιουργήθηκε η νέα διαδικασία, η οποία αφορά στο super market «AB Βασιλόπουλος». Στο κριτήριο support (βλ. παρ 3.7.2). διατηρήθηκε η προκαθορισμένη τιμή 0,95 αλλά και στην τιμή του κριτηρίου lift στον operator διατηρήθηκε η προκαθορισμένη τιμή του συστήματος. Τα 10 πρώτα αποτελέσματα σε σύνολο 136 αποτελεσμάτων που προέκυψαν από την εύρεση των συχνά εμφανιζόμενων προϊόντων ή συνόλων προϊόντων εμφανίζονται στον πίνακα 4-7.

Παρατηρείται ωστόσο, πως τα αποτελέσματα για το καλύτερο δυνατό support δεν είναι πάνω από 0,95, όπως ζητήθηκε. Αυτό συμβαίνει, διότι επιλέχθηκε στις παραμέτρους του FP Growth το «find min number of datasets», το οποίο μειώνει την τιμή του κριτηρίου support, έως ότου αυτό εμφανίσει αποτελέσματα. Δοκιμάστηκε αρχικά χωρίς να επιλεγεί όμως δεν έβγαλε κανένα κανόνα συσχέτισης, καθώς όπως φαίνεται στον πίνακα 4-7, το support είναι κάτω από 0,95.

Πίνακας 4-7: Παρουσίαση των 10 πρώτων πιο συχνά εμφανιζόμενων προϊόντων

Support	Κατηγορία Προϊόντων 1
0.226	Αλλαντικά και Delicatessen
0.226	Πρωϊνό, ροφήματα και σνακ
0.221	Ποτά, χυμοί, αναψυκτικά
0.216	Συσκευασμένα τρόφιμα
0.216	Φρούτα και λαχανικά
0.211	Κατεψυγμένα τρόφιμα
0.201	Βιολογικά – Vegan προϊόντα
0.201	Οικιακή φροντίδα (καθαριστικά, χαρτικά)
0.196	Γαλακτομικά και Είδη ψυγείου
0.186	Τυροκομικά

Παρατηρείται ότι αλλαντικά και delicatessen προϊόντα καθώς και είδη πρωινού, ροφήματα και σνακ αγοράζονται από το 22,6% των καταναλωτών που επισκέπτονται το super market της εταιρίας ΑΒ Βασιλόπουλος, εφόσον ο δείκτης support είναι 0,226 ενώ σε ποσοστό 22,1% αγοράζονται ποτά, χυμοί και αναψυκτικά.

Στον πίνακα 4-8 παρουσιάζονται οι κανόνες συσχέτισης που εξήχθησαν από τα παραπάνω δεδομένα με βάση τον δείκτη support. Ταξινομήθηκαν από τον μεγαλύτερο στον μικρότερο και παρουσιάζονται ενδεικτικά 10 κανόνες σε σύνολο 470 κανόνων.

Πίνακας 4-8: Παρουσίαση των κανόνων συσχέτισης

Κατηγορία Προϊόντων 1	Κατηγορία Προϊόντων 2	Support	Confidence	Lift
-----------------------	-----------------------	---------	------------	------

Πρωινό, ροφήματα και σνακ	Κατεψυγμένα τρόφιμα	0.181	0.8	3.79
Ποτά, χυμοί, αναψυκτικά	Κατεψυγμένα τρόφιμα	0.181	0.818	3.877
Κατεψυγμένα τρόφιμα	Πρωινό, ροφήματα και σνακ	0.181	0.857	3.79
Κατεψυγμένα τρόφιμα	Ποτά, χυμοί, αναψυκτικά	0.181	0.857	3.877
Αλλαντικά και Delicatessen	Πρωινό, ροφήματα και σνακ	0.176	0.778	3.44
Πρωινό, ροφήματα και σνακ	Αλλαντικά και Delicatessen	0.176	0.778	3.44
Πρωινό, ροφήματα και σνακ	Συσκευασμένα τρόφιμα	0.176	0.778	3.599
Συσκευασμένα τρόφιμα	Πρωινό, ροφήματα και σνακ	0.176	0.814	3.599
Πρωινό, ροφήματα και σνακ	Ποτά, χυμοί, αναψυκτικά	0.171	0.756	3.417
Ποτά, χυμοί, αναψυκτικά	Πρωινό, ροφήματα και σνακ	0.171	0.773	3.417

Τα αποτελέσματα που εξήχθησαν από τους κανόνες συσχέτισης δείχνουν ότι με βάση το κριτήριο της εμπιστοσύνης, οι καταναλωτές που αγοράζουν κατεψυγμένα τρόφιμα θα αγοράσουν με πολύ μεγάλη πιθανότητα και προϊόντα πρωινού, ροφήματα και σνακ. Αυτό γίνεται αντιληπτό καθώς το κριτήριο confidence παίρνει την τιμή 0,857. Με την ίδια πιθανότητα όσοι καταναλωτές αγοράσουν κατεψυγμένα τρόφιμα θα αγοράσουν

επίσης ποτά, χυμούς και αναψυκτικά. Επιπρόσθετα, το κριτήριο Lift σε ένα μεγάλο ποσοστό των κανόνων συσχέτισης είναι πάνω από 3 και σε ορισμένους κανόνες συσχέτισης φτάνει μέχρι και το 5, γεγονός που καταδεικνύει αφενός, ότι στους συγκεκριμένους κανόνες η αγορά του δεύτερου προϊόντος επηρεάζεται θετικά από την αγορά του πρώτου ή του πρώτου συνόλου προϊόντων και αφετέρου ότι δεν στηρίζεται σε ένα τυχαίο γεγονός.

4.6.3. Αποτελέσματα Market Basket Analysis για το Super Market «Σκλαβενίτης»

Αφού προηγήθηκαν οι παραπάνω επεξεργασίες εισάχθηκαν αρχικά τα δεδομένα για τις αγορές των καταναλωτών από το super market «Σκλαβενίτης». Για την τιμή του δείκτη support και του κριτηρίου lift για τον operator “Create Association Rules” χρησιμοποιήθηκε η δοσμένη από το σύστημα τιμή.

Τα 10 πρώτα αποτελέσματα σε σύνολο 32 αποτελεσμάτων που προέκυψαν από την εύρεση των συχνά εμφανιζόμενων προϊόντων ή συνόλων προϊόντων παρατίθενται στον πίνακα 4-9.

Παρατηρείται ωστόσο, πως τα αποτελέσματα για το καλύτερο δυνατό support δεν είναι πάνω από 0,95 όπως ζητήθηκε. Αυτό συμβαίνει, διότι επιλέχθηκε στις παραμέτρους του FP Growth το «find min number of datasets», το οποίο μειώνει την τιμή του κριτηρίου support έως ότου αυτό εμφανίσει αποτελέσματα. Δοκιμάστηκε αρχικά χωρίς να επιλεγεί, όμως δεν έβγαλε κανένα κανόνα συσχέτισης, καθώς όπως φαίνεται στον πίνακα 4-9, το support είναι κάτω από 0,95.

Πίνακας 4-9: Παρουσίαση των 10 πρώτων πιο συχνά εμφανιζόμενων προϊόντων

Support	Κατηγορία Προϊόντων 1	Κατηγορία Προϊόντων 2
0.447	Φρούτα και λαχανικά	
0.417	Γαλακτομικά και Είδη ψυγείου	
0.412	Πρωινό, ροφήματα και σνακ	
0.397	Συσκευασμένα τρόφιμα	
0.387	Οικιακή φροντίδα (καθαριστικά, χαρτικά)	
0.382	Ποτά, χυμοί, αναψυκτικά	
0.377	Τυροκομικά	

0.357	Φρούτα και λαχανικά	Γαλακτομικά και Είδη ψυγείου
0.352	Φρούτα και λαχανικά	Πρωινό, ροφήματα και σνακ
0.352	Γαλακτομικά και Είδη ψυγείου	Πρωινό, ροφήματα και σνακ

Παρατηρείται ότι τα φρούτα και τα λαχανικά αγοράζονται από το 44,7% των καταναλωτών και τα γαλακτοκομικά και είδη ψυγείου αγοράζονται από το 41,7% των καταναλωτών που επισκέπτονται το super market «Σκλαβενίτης», ενώ ακολουθούν τα προϊόντα για πρωινό, ροφήματα και σνακ που αγοράζονται από το 41,2% των καταναλωτών, καθώς η τιμή του support αντίστοιχα για τα παραπάνω είναι 0,447, 0,417 και 0,412.

Η εξαγωγή των κανόνων συσχέτισης των παραπάνω συνόλων προϊόντων παρουσιάζεται στον πίνακα 4-10. Παρουσιάζονται οι 10 πρώτοι κανόνες συσχέτισης σε σύνολο 490 κανόνων, αφού ταξινομήθηκαν σε φθίνουσα σειρά με βάση τον δείκτη support.

Πίνακας 4-10: Παρουσίαση των κανόνων συσχέτισης

Κατηγορία Προϊόντων 1	Κατηγορία Προϊόντων 2	Support	Confidence	Lift
Φρούτα και λαχανικά	Γαλακτομικά και Είδη ψυγείου	0.357	0.798	1.913
Γαλακτομικά και Είδη ψυγείου	Φρούτα και λαχανικά	0.357	0.855	1.913
Φρούτα και λαχανικά	Πρωινό, ροφήματα και σνακ	0.352	0.787	1.909
Γαλακτομικά και Είδη ψυγείου	Πρωινό, ροφήματα και σνακ	0.352	0.843	2.047
Πρωινό, ροφήματα και σνακ	Φρούτα και λαχανικά	0.352	0.854	1.909

Πρωινό, ροφήματα και σνακ	Γαλακτομικά και Είδη ψυγείου	0.352	0.854	2.047
Γαλακτομικά και Είδη ψυγείου	Τυροκομικά	0.337	0.807	2.142
Πρωινό, ροφήματα και σνακ	Συσκευασμένα τρόφιμα	0.337	0.817	2.058
Συσκευασμένα τρόφιμα	Πρωινό, ροφήματα και σνακ	0.337	0.848	2.058
Συσκευασμένα τρόφιμα	Ποτά, χυμοί, αναψυκτικά	0.337	0.848	2.221

Ο δείκτης lift είναι αρκετά ικανοποιητικός εφόσον η τιμή του είναι πάνω από 1. Επίσης, ο δείκτης εμπιστοσύνης είναι αρκετά υψηλός, το οποίο σημαίνει ότι για παράδειγμα για κάποιον που αγοράζει γαλακτοκομικά και είδη ψυγείου υπάρχει πιθανότητα 85,50% να αγοράσει και φρούτα και λαχανικά από το συγκεκριμένο super market, εφόσον το confidence έχει την τιμή 0,855. Όπως επίσης και με το κριτήριο support γίνεται αντιληπτό ότι τα παραπάνω δύο είδη βρίσκονται ταυτόχρονα σε συχνότητα 35,7% του συνόλου των προϊόντων που αγοράζονται από το συγκεκριμένο super market.

4.6.4. Αποτελέσματα Market Basket Analysis για το Super Market «Χαλκιάδακης»

Για την εξαγωγή των συχνά αγοραζόμενων προϊόντων ή σετ προϊόντων καθώς και των κανόνων συσχέτισης, χρησιμοποιήθηκαν οι ίδιες τιμές κριτηρίων όπως και στην επεξεργασία των δεδομένων για τα προηγούμενα δύο super market, δηλαδή στην τιμή του κριτηρίου support και στο lift διατηρήθηκε η δοσμένη από το σύστημα τιμή.

Στον πίνακα 4-11 παρουσιάζονται ενδεικτικά τα 10 πρώτα αποτελέσματα σε σύνολο 204 αποτελεσμάτων που προέκυψαν από την εύρεση των συχνά εμφανιζόμενων προϊόντων ή συνόλων προϊόντων.

Παρατηρείται ωστόσο, πως τα αποτελέσματα για το καλύτερο δυνατό support δεν είναι πάνω από 0,95, όπως ζητήθηκε. Αυτό συμβαίνει, διότι επιλέχθηκε στις παραμέτρους του FP Growth το «find min number of datasets», το οποίο μειώνει την τιμή του κριτηρίου support έως ότου αυτό εμφανίσει αποτελέσματα. Δοκιμάστηκε αρχικά χωρίς να επιλεγεί όμως δεν έβγαλε κανένα κανόνα συσχέτισης, καθώς όπως φαίνεται στον πίνακα 4-11, το support είναι κάτω από 0,95.

Πίνακας 4-11: Παρουσίαση των 10 πρώτων πιο συχνά εμφανιζόμενων προϊόντων

Support	Κατηγορία Προϊόντων 1	Κατηγορία Προϊόντων 2
0.276	Κρητικά προϊόντα	
0.251	Τυροκομικά	
0.246	Γαλακτομικά και Είδη ψυγείου	
0.226	Πρωινό, ροφήματα και σνακ	
0.221	Τυροκομικά	Γαλακτομικά και Είδη ψυγείου
0.211	Ποτά, χυμοί, αναψυκτικά	
0.206	Σπίτι και κουζίνα	
0.206	Συσκευασμένα τρόφιμα	
0.201	Φρούτα και λαχανικά	
0.196	Αλλαντικά και Delicatessen	

Στον παραπάνω πίνακα (Πίνακας 4-11) παρατηρείται ότι το ποσοστό των καταναλωτών που επισκέπτεται το super market «Χαλκιάδκης» και αγοράζει κρητικά προϊόντα ανέρχεται στο 27,6% και για τυροκομικά το 25,1%. Ενώ ταυτόχρονα τυροκομικά και γαλακτοκομικά και είδη ψυγείου αγοράζει το 22,1%. Αυτό γίνεται αντιληπτό, καθώς ο δείκτης support είναι 0,276, 0,251 και 0,221 για κάθε περίπτωση.

Στον πίνακα 4-12 παρουσιάζονται οι κανόνες συσχέτισης των παραπάνω προϊόντων. Παρουσιάζονται ενδεικτικά οι 10 πρώτοι κανόνες συσχέτισης σε σύνολο 1133 κανόνων.

Πίνακας 4-12: Παρουσίαση των κανόνων συσχέτισης

Κατηγορία Προϊόντων 1	Κατηγορία Προϊόντων 2	Support	Confidence	Lift
Τυροκομικά	Γαλακτομικά και Είδη ψυγείου	0.221	0.88	3.574
Γαλακτομικά και Είδη ψυγείου	Τυροκομικά	0.221	0.898	3.574
Κρητικά προϊόντα	Γαλακτομικά και Είδη ψυγείου	0.191	0.691	2.806
Γαλακτομικά και Είδη ψυγείου	Κρητικά προϊόντα	0.191	0.776	2.806
Γαλακτομικά και Είδη ψυγείου	Πρωινό, ροφήματα και σνακ	0.191	0.776	3.429
Πρωινό, ροφήματα και σνακ	Γαλακτομικά και Είδη ψυγείου	0.191	0.844	3.429
Πρωινό, ροφήματα και σνακ	Ποτά, χυμοί, αναψυκτικά	0.191	0.844	4.001
Ποτά, χυμοί, αναψυκτικά	Πρωινό, ροφήματα και σνακ	0.191	0.905	4.001
Ποτά, χυμοί, αναψυκτικά	Συσκευασμένα τρόφιμα	0.191	0.905	4.391
Συσκευασμένα τρόφιμα	Ποτά, χυμοί, αναψυκτικά	0.191	0.927	4.391

Ο δείκτης lift είναι αρκετά ικανοποιητικός εφόσον η τιμή του είναι πάνω από 1 και σε πολλές περιπτώσεις φτάνει και το 5 (Παράρτημα). Επίσης, ο δείκτης εμπιστοσύνης είναι αρκετά υψηλός, το οποίο σημαίνει ότι για παράδειγμα για κάποιον που αγοράζει συσκευασμένα τρόφιμα, υπάρχει πιθανότητα 92,7% να αγοράσει και ποτά, χυμούς και αναψυκτικά από το συγκεκριμένο super market. Με το κριτήριο support γίνεται αντιληπτό ότι τα παραπάνω τρία είδη βρίσκονται ταυτόχρονα σε συχνότητα 19,1% του συνόλου των προϊόντων που αγοράζονται από το συγκεκριμένο super market.

4.6.5. Αποτελέσματα Market Basket Analysis για το Super Market «LIDL»

Για το super market «LIDL» ακολουθήθηκε ακριβώς η ίδια διαδικασία με τα προηγούμενα super markets.

Τα αποτελέσματα για τα πιο συχνά αγοραζόμενα προϊόντα ή συνδυασμό αγοραζόμενων προϊόντων εμφανίζονται στον πίνακα 4-13. Παρουσιάζονται ενδεικτικά τα πρώτα 10 αποτελέσματα σε σύνολο 109 αποτελεσμάτων.

Παρατηρείται ωστόσο και εδώ, πως τα αποτελέσματα για το καλύτερο δυνατό support δεν είναι πάνω από 0,95, όπως ζητήθηκε. Αυτό συμβαίνει, διότι επιλέχθηκε στις παραμέτρους του FP Growth το «find min number of datasets», το οποίο μειώνει την τιμή του κριτηρίου support, έως ότου αυτό εμφανίσει αποτελέσματα. Δοκιμάστηκε αρχικά χωρίς να επιλεγεί όμως δεν έβγαλε κανένα κανόνα συσχέτισης, καθώς όπως φαίνεται στον πίνακα 4-13, το support είναι κάτω από 0,95.

Πίνακας 4-13: Παρουσίαση των 10 πρώτων πιο συχνά εμφανιζόμενων προϊόντων

Support	Κατηγορία Προϊόντων 1	Κατηγορία Προϊόντων 2
0.322	Οικιακή φροντίδα (καθαριστικά, χαρτικά)	
0.291	Σπίτι και κουζίνα	
0.276	Εποχιακά είδη	
0.256	Ποτά, χυμοί, αναψυκτικά	
0.251	Οικιακή φροντίδα (καθαριστικά, χαρτικά)	Σπίτι και κουζίνα
0.241	Συσκευασμένα τρόφιμα	
0.236	Κατεψυγμένα τρόφιμα	

0.226	Αρτοζαχαροπλασείο	
0.186	Πρωινό, ροφήματα και σνακ	
0.186	Οικιακή φροντίδα (καθαριστικά, χαρτικά)	Ποτά, χυμοί, αναψυκτικά

Στον παραπάνω πίνακα (Πίνακας 4-13) φαίνεται ότι το μεγαλύτερο ποσοστό των καταναλωτών προτιμάει το super market «LIDL» για προϊόντα οικιακής φροντίδας, όπως π.χ. καθαριστικά και χαρτικά. Συγκεκριμένα, αφορά το 32,2% των καταναλωτών και ακολουθεί το 29,1% των καταναλωτών που θα προμηθευτεί είδη σπιτιού και κουζίνας.

Στον πίνακα 4-14 παρουσιάζονται ενδεικτικά οι 10 πρώτοι κανόνες συσχέτισης που προκύπτουν από τα παραπάνω προϊόντα.

Πίνακας 4-14: Παρουσίαση κανόνων συσχέτισης

Κατηγορία Προϊόντων 1	Κατηγορία Προϊόντων 2	Support	Confidence	Lift
Οικιακή φροντίδα (καθαριστικά, χαρτικά)	Σπίτι και κουζίνα	0.251	0.781	2.68
Σπίτι και κουζίνα	Οικιακή φροντίδα (καθαριστικά, χαρτικά)	0.251	0.862	2.68
Οικιακή φροντίδα (καθαριστικά, χαρτικά)	Ποτά, χυμοί, αναψυκτικά	0.186	0.578	2.256
Σπίτι και κουζίνα	Εποχιακά είδη	0.186	0.638	2.308
Εποχιακά είδη	Σπίτι και κουζίνα	0.186	0.673	2.308
Ποτά, χυμοί, αναψυκτικά	Οικιακή φροντίδα	0.186	0.725	2.256

	(καθαριστικά, χαρτικά)			
Ποτά, χυμοί, αναψυκτικά	Συσκευασμένα τρόφιμα	0.181	0.706	2.926
Συσκευασμένα τρόφιμα	Ποτά, χυμοί, αναψυκτικά	0.181	0.75	2.926
Συσκευασμένα τρόφιμα	Κατεψυγμένα τρόφιμα	0.181	0.75	3.176
Κατεψυγμένα τρόφιμα	Συσκευασμένα τρόφιμα	0.181	0.766	3.176

Για το συγκεκριμένο super market, όπως φαίνεται και από τον παραπάνω πίνακα (Πίνακας 4-14), το ποσοστό των πελατών είναι αρκετά μικρό κρίνοντας από την τιμή του κριτηρίου support, ωστόσο τα συμπεράσματα που εξάγονται για τους κανόνες συσχέτισης των προϊόντων είναι αρκετά ικανοποιητικά. Πιο συγκεκριμένα, οι πελάτες που θα προμηθευτούν είδη σπιτιού και κουζίνας είναι πιθανό σε ποσοστό 86,2% να προμηθευτούν και προϊόντα οικιακής φροντίδας. Επίσης, και ο δείκτης Lift είναι ικανοποιητικός καθώς στο μεγαλύτερο μέρος των κανόνων συσχέτισης κυμαίνεται μεταξύ του 2 και 3.

4.6.6. Αποτελέσματα Market Basket Analysis για το Super Market «SYNKA»

Για το super market «SYNKA» ακολουθήθηκε ακριβώς η ίδια διαδικασία με τα προηγούμενα super markets.

Τα αποτελέσματα για τα πιο συχνά αγοραζόμενα προϊόντα ή συνδυασμό αγοραζόμενων προϊόντων εμφανίζονται στον πίνακα 4-16. Παρουσιάζονται ενδεικτικά τα πρώτα 10 αποτελέσματα σε σύνολο 137 αποτελεσμάτων.

Παρατηρείται ωστόσο, πως τα αποτελέσματα για το καλύτερο δυνατό support δεν είναι πάνω από 0,95, όπως ζητήθηκε. Αυτό συμβαίνει, διότι επιλέχθηκε στις παραμέτρους του FP Growth το «find min number of datasets», το οποίο μειώνει την τιμή του κριτηρίου support έως ότου αυτό εμφανίσει αποτελέσματα. Δοκιμάστηκε αρχικά χωρίς να επιλεγεί όμως δεν έβγαλε κανένα κανόνα συσχέτισης, καθώς όπως φαίνεται στον πίνακα 4-16, το support είναι κάτω από 0,95.

Πίνακας 4-15: Παρουσίαση των πιο συχνά εμφανιζόμενων προϊόντων

Support	Κατηγορία Προϊόντων 1	Κατηγορία Προϊόντων 2
0.492	Κρητικά προϊόντα	
0.422	Ποτά, χυμοί, αναψυκτικά	
0.402	Γαλακτομικά και Είδη ψυγείου	
0.402	Τυροκομικά	
0.392	Οικιακή φροντίδα (καθαριστικά, χαρτικά)	
0.372	Πρωινό, ροφήματα και σνακ	
0.372	Συσκευασμένα τρόφιμα	
0.352	Κατεψυγμένα τρόφιμα	
0.342	Κρητικά προϊόντα	Ποτά, χυμοί, αναψυκτικά
0.342	Ποτά, χυμοί, αναψυκτικά	Γαλακτομικά και Είδη ψυγείου

Όπως φαίνεται στον παραπάνω πίνακα (Πίνακας 4-17), οι καταναλωτές που επισκέπτονται τα super markets «SYNKA» σε ποσοστό 49,2% θα προμηθευτούν κρητικά προϊόντα. Ένω, ένα επίσης μεγάλο ποσοστό 42,2% θα προμηθευτεί ποτά, χυμούς, αναψυκτικά.

Παρακάτω παρουσιάζονται ενδεικτικά οι 10 πρώτοι κανόνες συσχέτισης που προέκυψαν από τα παραπάνω προϊόντα.

Πίνακας 4-16: Παρουσίαση κανόνων συσχέτισης

Κατηγορία Προϊόντων 1	Κατηγορία Προϊόντων 2	Support	Confidence	Lift
Κρητικά προϊόντα	Ποτά, χυμοί, αναψυκτικά	0.342	0.694	1.644
Ποτά, χυμοί, αναψυκτικά	Κρητικά προϊόντα	0.342	0.810	1.644

Ποτά, χυμοί, αναψυκτικά	Γαλακτομικά και Είδη ψυγείου	0.342	0.810	2.014
Ποτά, χυμοί, αναψυκτικά	Συσκευασμένα τρόφιμα	0.342	0.810	2.177
Γαλακτομικά και Είδη ψυγείου	Ποτά, χυμοί, αναψυκτικά	0.342	0.850	2.014
Συσκευασμένα τρόφιμα	Ποτά, χυμοί, αναψυκτικά	0.342	0.919	2.177
Κρητικά προϊόντα	Γαλακτομικά και Είδη ψυγείου	0.332	0.673	1.675
Κρητικά προϊόντα	Τυροκομικά	0.332	0.673	1.675
Ποτά, χυμοί, αναψυκτικά	Οικιακή φροντίδα (καθαριστικά, χαρτικά)	0.332	0.786	2.005
Γαλακτομικά και Είδη ψυγείου	Κρητικά προϊόντα	0.332	0.825	1.675

Τα αποτελέσματα που εξήχθησαν από του κανόνες συσχέτισης δείχνουν ότι με βάση το κριτήριο της εμπιστοσύνης, οι καταναλωτές που αγοράζουν συσκευασμένα τρόφιμα θα αγοράσουν με πολύ μεγάλη πιθανότητα και ποτά, χυμούς και αναψυκτικά. Αυτό γίνεται αντιληπτό καθώς το κριτήριο confidence παίρνει την τιμή 0,919, δηλαδή αυτό θα συμβεί με πιθανότητα 91,9%. Επίσης, και ο δείκτης Lift είναι ικανοποιητικός καθώς στο μεγαλύτερο μέρος των κανόνων συσχέτισης κυμαίνεται μεταξύ του 2 και 3, το οποίο δείχνει ότι αυτό δεν αποτελεί ένα τυχαίο γεγονός καθώς αυτός ο δείκτης είναι πάνω από 1.

4.6.7. Αποτελέσματα Market Basket Analysis για την επιλογή «Κανένα»

Στο ερωτηματολόγιο που μοιράστηκε εκτός από τα προαναφερθέντα super market υπήρχε και η επιλογή «Κανένα» σε περίπτωση που οι καταναλωτές επέλεγουν να προμηθεύονται προϊόντα κάποιων κατηγοριών από άλλα καταστήματα, όπως κρεοπωλεία, ιχθυοπωλεία, οπωροπωλεία και άλλα.

Στον πίνακα 4-18 παρατίθενται ενδεικτικά τα 10 πρώτα πιο συχνά εμφανιζόμενα προϊόντα για την επιλογή «Κανένα», δηλαδή τα πιο συνήθη προϊόντα που οι καταναλωτές προμηθεύονται από άλλου είδους καταστήματα εκτός των εναλλακτικών που δόθηκαν.

Παρατηρείται ωστόσο, πως τα αποτελέσματα για το καλύτερο δυνατό support δεν είναι πάνω από 0,95, όπως ζητήθηκε. Αυτό συμβαίνει, διότι επιλέχθηκε στις παραμέτρους του FP Growth το «find min number of datasets», το οποίο μειώνει την τιμή του κριτηρίου support έως ότου αυτό εμφανίσει αποτελέσματα. Δοκιμάστηκε αρχικά χωρίς να επιλεγεί όμως δεν έβγαλε κανένα κανόνα συσχέτισης, καθώς όπως φαίνεται στον πίνακα 4-18, το support είναι κάτω από 0,95.

Πίνακας 4-17: Παρουσίαση των πιο συχνά εμφανιζόμενων προϊόντων

Support	Κατηγορία Προϊόντων 1	Κατηγορία Προϊόντων 2
0.503	Φρέσκο Ψάρι	
0.467	Κατοικίδια	
0.407	Βρεφικά προϊόντα	
0.352	Φρέσκο Κρέας	
0.337	Βιολογικά – Vegan προϊόντα	
0.332	Φρέσκο Ψάρι	Κατοικίδια
0.317	Φρέσκο Ψάρι	Φρέσκο Κρέας
0.307	Εποχιακά είδη	
0.296	Φρέσκο Ψάρι	Βρεφικά προϊόντα
0.291	Αρτοζαχαροπλαστείο	

Όπως γίνεται αντιληπτό, το 50,3% των ερωτηθέντων προμηθεύεται φρέσκο ψάρι από άλλα καταστήματα και σύμφωνα με τα σχόλια που δόθηκαν στο ερωτηματολόγιο, προτιμάει τα ιχθυοπωλεία. Ένα επίσης μεγάλο ποσοστό της τάξης του 46,7% επιλέγει διαφορετικά καταστήματα και για τα προϊόντα που αφορούν τα κατοικίδια.

Στον πίνακα 4-19 παρουσιάζονται οι κανόνες συσχέτισης ανάμεσα στα προϊόντα που δεν αγοράζονται από τους ερωτηθέντες από κανένα από τα παραπάνω super markets.

Πίνακας 4-18: Παρουσίαση κανόνων συσχέτισης

Κατηγορία Προϊόντων 1	Κατηγορία Προϊόντων 2	Support	Confidence	Lift
Φρέσκο Ψάρι	Κατοικίδια	0.332	0.66	1.412
Κατοικίδια	Φρέσκο Ψάρι	0.332	0.710	1.412
Φρέσκο Ψάρι	Φρέσκο Κρέας	0.317	0.63	1.791
Φρέσκο Κρέας	Φρέσκο Ψάρι	0.317	0.9	1.791
Φρέσκο Ψάρι	Βρεφικά προϊόντα	0.296	0.59	1.450
Βρεφικά προϊόντα	Φρέσκο Ψάρι	0.296	0.729	1.450
Κατοικίδια	Βρεφικά προϊόντα	0.286	0.613	1.506
Βρεφικά προϊόντα	Κατοικίδια	0.286	0.704	1.506
Φρέσκο Ψάρι	Βιολογικά – Vegan προϊόντα	0.246	0.49	1.455
Κατοικίδια	Βιολογικά – Vegan προϊόντα	0.246	0.527	1.565

Οι κανόνες συσχέτισης που παρουσιάζονται παραπάνω είναι ενδεικτικά οι πρώτοι 10 σε σύνολο 630 κανόνων, όπως εξάχθηκαν από το RapidMiner, σε φθίνουσα σειρά με βάση το κριτήριο support. Για τους καταναλωτές που επιλέγουν να προμηθευτούν από διαφορετικά καταστήματα φρέσκο κρέας είναι πιθανό κατά 90% να προμηθευτούν και φρέσκο ψάρι. Ωστόσο ο συνδυασμός αυτός εμφανίζεται σε ποσοστό 31,7%. Επιπλέον, το κριτήριο lift, είναι μεγαλύτερο του 1, το οποίο είναι αρκετά ικανοποιητικό, όμως συγκριτικά με τα προηγούμενα αποτελέσματα δεν είναι τόσο μεγάλο. Το κριτήριο lift κυμαίνεται σε τιμές άνω του 1 και μέχρι 3 για το σύνολο των κανόνων συσχέτισης.

Συμπεράσματα MBA

Συνοψίζοντας, στο Market Basket Analysis εξήχθησαν τα παρακάτω συμπεράσματα. Αφενός το super market που παρουσίασε το μεγαλύτερο support ήταν αυτό της εταιρίας «Σκλαβενίτης», το οποίο ήταν αναμενόμενο, καθώς και από τα αποτελέσματα της μεθόδου UTASTAR, γίνεται αντιληπτό ότι είναι το πλέον προτιμώμενο super market με την μεγαλύτερη τιμή του support να είναι στο 0,357 για συνδυασμό προϊόντων. Συνοπτικά, οι δείκτες support, confidence και lift για κάθε super market, αλλά και για την επιλογή άλλου καταστήματος («Κανένα») παρουσιάζονται παρακάτω με την ταξινόμηση να γίνεται με βάση τον μεγαλύτερο δείκτη support.

Πίνακας 4-20: Συνοπτική παρουσίαση κανόνων συσχέτισης για όλα τα super market

Super market	Κατηγορία Προϊόντων 1	Κατηγορία Προϊόντων 2	Support	Confidence	Lift
AB Βασιλόπουλος	Πρωινό, ροφήματα και σνακ	Κατεψυγμένα τρόφιμα	0,181	0,8	3,79
Σκλαβενίτης	Φρούτα και λαχανικά	Γαλακτομικά και Είδη ψυγείου	0,357	0,798	1,913
Χαλκιαδάκης	Τυροκομικά	Γαλακτοκομικά και Είδη ψυγείου	0,221	0,880	3,574
LIDL	Οικιακή φροντίδα	Σπίτι και κουζίνα	0,251	0,781	2,68
SYNKA	Κρητικά προϊόντα	Ποτά, χυμοί, αναψυκτικά	0,342	0,694	1,644
Κανένα	Φρέσκο ψάρι	Κατοικίδια	0,332	0,660	1,412

Συνολικά, το κριτήριο του support δεν είναι μεγαλύτερο του 0.6 σε καμία από τις παραπάνω επιλογές. Ωστόσο, τα κριτήρια confidence και lift έχουν αρκετά υψηλές τιμές σχεδόν σε όλες τις επιλογές. Αυτό επιβεβαιώνει το γεγονός ότι οι συσχετίσεις που γίνονται δεν αποτελούν τυχαία γεγονότα, όσον αφορά στο κριτήριο lift και επιπλέον ότι υπάρχει υψηλό κριτήριο εμπιστοσύνης ανάμεσα στις συσχετίσεις των προϊόντων που αγοράζονται συνδυαστικά από ένα super market.

Κεφάλαιο 5 ΣΥΜΠΕΡΑΣΜΑΤΑ – ΠΡΟΤΑΣΕΙΣ

Συνοψίζοντας, στη παρούσα διπλωματική εργασία αρχικά δημιουργήθηκε το ερωτηματολόγιο και από τις απαντήσεις που δόθηκαν, δημιουργήθηκε η βάση δεδομένων που χρησιμοποιήθηκε στην επεξεργασία. Αρχικά, η επεξεργασία δεδομένων περιελάμβανε την στατιστική τους ανάλυση. Στη συνέχεια, πραγματοποιήθηκε η προεπεξεργασία των δεδομένων. Σε δεύτερο στάδιο για την εξαγωγή των βαρών των κριτηρίων και των ολικών χρησιμότητων από τον πολυκριτήριο πίνακα του ερωτηματολογίου, χρησιμοποιήθηκε ο αλγόριθμος πολυκριτήριας ανάλυσης UTASTAR μέσω του συστήματος MARKEK. Έπειτα, πραγματοποιήθηκε συσταδοποίηση των βαρών των κριτηρίων και των ολικών χρησιμότητων με χρήση του αλγορίθμου k-means σε γλώσσα προγραμματισμού MATLAB. Τέλος, χρησιμοποιήθηκε η μέθοδος Market Basket Analysis για την εξαγωγή κανόνων συσχέτισης των προϊόντων που αγοράζονται από τις διαθέσιμες εναλλακτικές επιλογές.

Εφαρμόζοντας την παραπάνω μεθοδολογία προκύπτουν συμπεράσματα για το προφίλ των καταναλωτών, την καταναλωτική συμπεριφορά τους, καθώς και τα κριτήρια με τα οποία πραγματοποιούν τις αγορές τους. Επίσης, εξάγονται συμπεράσματα για τον συνδυασμό των προϊόντων που αγοράζουν από κάθε εναλλακτική επιλογή super market.

Μετά από την συσταδοποίηση των καταναλωτών με βάση τα βάρη των κριτηρίων και τις ολικές χρησιμότητες διαπιστώθηκε ότι όσον αφορά στα βάρη των κριτηρίων, η πρώτη συστάδα καταναλωτών επέλεξε ως σημαντικότερο κριτήριο αυτό της τιμής των προϊόντων και στη συνέχεια αυτό των προσφορών. Η δεύτερη συστάδα καταναλωτών που δημιουργήθηκε, θεωρεί επίσης σημαντικότερο το κριτήριο της τιμής των προϊόντων με μεγάλη διαφορά από τα υπόλοιπα κριτήρια. Στην τρίτη συστάδα, οι καταναλωτές που ανήκουν σ' αυτήν, επέλεξαν ως σημαντικότερο κριτήριο το κριτήριο της απόστασης του super market από τον τόπο κατοικίας τους και στη συνέχεια το κριτήριο της τιμής των προϊόντων. Όσον αφορά στις ολικές χρησιμότητες η πρώτη συστάδα επέλεξε το super market της εταιρίας «Σκλαβενίτης» και αμέσως επόμενο της εταιρίας «SYNKA». Στην δεύτερη συστάδα οι καταναλωτές επέλεξαν το super market «SYNKA», ενώ στην τρίτη συστάδα οι καταναλωτές επέλεξαν ως προτιμώτερο super market το «Χαλκιάδακης».

Στο κομμάτι του Market Basket Analysis τα αποτελέσματα που εξήχθησαν αφορούσαν στην συχνότητα που εμφανίζονται κάποια/ες κατηγορίες προϊόντων στο καλάθι των καταναλωτών ξεχωριστά ανά super market, καθώς και οι κανόνες συσχέτισης των κατηγοριών μεταξύ τους. Χρησιμοποιήθηκαν τρία κριτήρια για την αξιολόγηση των κανόνων συσχέτισης, τα οποία ήταν τα: support (υποστήριξη), confidence (εμπιστοσύνη) και lift (κατώφλι). Τα κριτήρια confidence και lift ήταν ικανοποιητικά για

το μεγαλύτερο μέρος των κανόνων συσχέτισης, ενώ και το κριτήριο support πήρε αρκετά ικανοποιητικές τιμές σε πολλές περιπτώσεις. Συγκεκριμένα για κάθε super market θα αναφερθεί ενδεικτικά ο πρώτος κανόνας συσχέτισης ο οποίος έχει την μεγαλύτερη τιμή του κριτηρίου confidence.

Πίνακας 5-1: Συνοπτική παρουσίαση κανόνων συσχέτισης

Super market	Κατηγορία Προϊόντων 1	Κατηγορία Προϊόντων 2	Support	Confidence	Lift
AB Βασιλόπουλος	Κατεψυγμένα τρόφιμα, Αντισηπτική προστασία	Πρωινό, ροφήματα και σνακ	0,1407	1	4,422
Σκλαβενίτης	Συσκευασμένα τρόφιμα, Τυροκομικά	Γαλακτομικά και Είδη ψυγείου	0,2764	0,9649	2,313
Χαλκιαδάκης	Ποτά, χυμοί, αναψυκτικά, Υγεία και ομορφιά	Πρωινό, ροφήματα και σνακ	0.1407	1	4,422
LIDL	Σπίτι και κουζίνα, Ποτά, χυμοί, αναψυκτικά, Συσκευασμένα τρόφιμα	Οικιακή φροντίδα (καθαριστικά, χαρτικά)	0.1256	0.9615	2,990
SYNKA	Οικιακή φροντίδα (καθαριστικά, χαρτικά), Πρωινό, ροφήματα και σνακ, Συσκευασμένα τρόφιμα	Ποτά, χυμοί, αναψυκτικά	0.2663	0.9815	2,325
Κανένα	Κατοικίδια, Βρεφικά προϊόντα, Φρέσκο Κρέας	Φρέσκο Ψάρι	0.1256	0.9259	1,843

Με όλες τις παραπάνω επεξεργασίες μπορούν να εξαχθούν συμπεράσματα για κάθε super market χωριστά προκειμένου να αυξηθεί το καταναλωτικό τους κοινό είτε βελτιώνοντας κάποια από τα κριτήρια με τα οποία οι καταναλωτές επιλέγουν να ψωνίσουν, είτε δημιουργώντας προωθητικές ενέργειες με βάση τις αγορές που πραγματοποιούν οι καταναλωτές, συνδυάζοντας διαφορετικά είδη, ανά super market.

ΔΙΕΘΝΗΣ ΒΙΒΛΙΟΓΡΑΦΙΑ

- Aiguo, Z., Ping, S., & Lanling, J. (2011). Application of Data Mining in Supermarket. IEEE. doi:10.1109/CECNET.2011.5769080
- Akpınar, M., Ozkan, B., Sayin, C., & Ceylan, R. (2010). Consumer risk perceptions towards food supply chain preferences: The risk of supermarket. *Journal of Food, Agriculture & Environment*, σσ. 256-260
- Chen, Y.-L., Tang, K., Shen, R.-J., & Hu, Y.-H. (2004). Market basket analysis in a multiple store environment. *Decision Support Systems*, σσ. 339-354. doi:10.1016/j.dss.2004.04.009
- Iakovou, S., Kanavos, A., & Tsakalidis, A. (2016). Customer Behaviour Analysis for Recommendation of Supermarket Ware. *Springer International Publishing Switzerland*, σσ. 471–480. doi:10.1007/978-3-319-44944-9_41
- Kadlaskar, A. (2021, October 2). *Analytics Vidhya*. Ανάκτηση από <https://www.analyticsvidhya.com/blog/2021/10/a-comprehensive-guide-on-market-basket-analysis/>
- Kaur, M., & Kang, S. (2016). Market Basket Analysis: Identify the changing trends of market data using association rule mining. *Procedia Computer Science*. doi:10.1016/j.procs.2016.05.180
- Li, S. (2017, September 25). *Towards Data Science*. Ανάκτηση από <https://towardsdatascience.com/a-gentle-introduction-on-market-basket-analysis-association-rules-fa4b986a40ce>
- Matsatsinis, N. F., & Siskos, Y. (2003). *Intelligent Support Systems for Marketing Decisions*. Springer Science & Business Media.
- Miteloudi, K. (2017). *Ανάπτυξη πολυκριτήριου συστήματος για τη λήψη ομαδικών αποφάσεων*. Chania
- Prithiviraj P. and Porkodi R., 2015. A Comparative Analysis of Association Rule Mining Algorithms in Data Mining: A study
- Raeder, T., & Chawla, N. (2010). Market basket analysis with networks. *Springer-Verlag*, σσ. 97-113. doi:10.1007/s13278-010-0003-7
- Siskos, Y., Grigoroudis, E., & Matsatsinis, N. F. (2016). "UTA Methods" in Multiple Criteria Decision Analysis. Στο J. R. Figueira, S. Greco, & M. Ehrgott, *Multiple Criteria Decision Analysis: State of the Art Surveys* (σσ. 315-362). Berlin: Springer Verlag. doi:10.1007/978-1-4939-3094-4_9
- Solgaard, H. S., & Hansen, T. (2003). A hierarchical Bayes model of choice between supermarket formats. *Journal of Retailing and Consumer Services*, σσ. 169-180. doi:10.1016/S0969-6989(03)00008-0

ΕΛΛΗΝΙΚΗ ΒΙΒΛΙΟΓΡΑΦΙΑ

- Ματσατσίνης, Ν. (1995). *Διδακτορική Διατριβή "Ένα Έμπειρο Σύστημα Υποστήριξης Αποφάσεων Μάρκετινγκ: Μεθοδολογία Υποστήριξης και Ολοκληρωμένη Αρχιτεκτονική"*. Χανιά: Πολυτεχνείο Κρήτης.
- Ματσατσίνης, Ν. (2010). *Συστήματα Υποστήριξης Αποφάσεων*. Αθήνα: Εκδόσεις Νέων Τεχνολογιών.
- Ματσατσίνης, Ν. (2021). *Επιχειρηματική Ευφυΐα, Επιχειρηματική Αναλυτική και Ανάλυση Μεγάλων Δεδομένων*. Εκδόσεις Νέων Τεχνολογιών.

<https://repository.kallipos.gr/bitstream/11419/1235/2/Kef. 8.pdf>

ΠΑΡΑΡΤΗΜΑ

Ερωτηματολόγιο



**ΠΟΛΥΤΕΧΝΕΙΟ
ΚΡΗΤΗΣ**

Σχολή Μηχανικών Παραγωγής & Διοίκησης

Έρευνα ανταγωνισμού-super market

Η παρούσα έρευνα πραγματοποιείται στο πλαίσιο εκπόνησης της διπλωματικής εργασίας μου στη Σχολή Μηχανικών Παραγωγής και Διοίκησης του Πολυτεχνείου Κρήτης. Σκοπός της διπλωματικής είναι η μελέτη και ανάλυση της συμπεριφοράς των καταναλωτών των super market με χρήση προχωρημένων μεθόδων αναλυτικής.

Το ερωτηματολόγιο είναι ΑΝΩΝΥΜΟ και η συμπλήρωσή του θα διαρκέσει λίγα λεπτά. Οι απαντήσεις θα χρησιμοποιηθούν αποκλειστικά και μόνο για εκπαιδευτικούς και ερευνητικούς σκοπούς.

Εάν έχετε οποιαδήποτε ερώτηση ή απορία στη συμπλήρωση του ερωτηματολογίου, μπορείτε να επικοινωνήσετε μαζί μου στο email: kkaditi@isc.tuc.gr

Σας ευχαριστώ για την βοήθειά σας

Με εκτίμηση

Καδίτη Κωνσταντίνα

Α. Δημογραφικά στοιχεία ερωτώμενου (Επιλέξτε το κελί που αντιστοιχεί στην απάντησή σας, καθώς και συμπληρώστε/γράψτε την απάντησή σας όπου χρειάζεται)

1) Φύλο:

- ☐ Γυναίκα
- ☐ Άνδρας

2) Ηλικία (σημειώστε την απάντησή σας (αριθμητικά)):

3) Μορφωτικό Επίπεδο:

- ☐ Απόφοιτος Δημοτικού
- ☐ Απόφοιτος Λυκείου
- ☐ Απόφοιτος Ι.Ε.Κ
- ☐ Φοιτητής Τριτοβάθμιας Εκπαίδευσης

- ☐ Πτυχιούχος Τριτοβάθμιας Εκπαίδευσης
- ☐ Κάτοχος Μεταπτυχιακού
- ☐ Κάτοχος Διδακτορικού
- ☐ Άλλο

4) Οικογενειακή κατάσταση:

- ☐ Έγγαμος
- ☐ Σε σχέση
- ☐ Ελεύθερος

5) Αριθμός τέκνων (σημειώστε την απάντησή σας (αριθμητικά)):

6) Από πόσα μέλη αποτελείται συνολικά η οικογένειά σας για τα οποία κάνετε αγορές από το super market (σημειώστε την απάντησή σας αριθμητικά):.....

7) Επάγγελμα:

- ☐ Αγρότης
- ☐ Άνεργος
- ☐ Αυτοαπασχολούμενος
- ☐ Δημόσιος Υπάλληλος
- ☐ Εισοδηματίας
- ☐ Ελεύθερος Επαγγελματίας
- ☐ Ιδιωτικός Υπάλληλος
- ☐ Μαθητής/Φοιτητής
- ☐ Οικιακά
- ☐ Συνταξιούχος
- ☐ Άλλο

8) Ετήσιο οικογενειακό εισόδημα:

- ☐ Έως 5.000€
- ☐ Από 5.001-10.000€
- ☐ Από 10.001-15.000€
- ☐ Από 15.001-25.000€
- ☐ Από 25.001-35.000€
- ☐ Από 35.001-45.000€
- ☐ Άνω των 45.000€

9) Θα σας παρακαλούσαμε να μας αναφέρετε σε ποια περιοχή των Χανίων κατοικείτε:

B. Αγορά προϊόντων από super market*(Επιλέξτε το κελί ή τα κελιά που*

αντιστοιχούν στις απαντήσεις σας, ανάλογα με το ζητούμενο)

1) Πόσο συχνά ψωνίζετε από super market:

- ☐ Καθημερινά
- ☐ 3-5 φορές την εβδομάδα
- ☐ 1-2 φορές την εβδομάδα
- ☐ 1-3 φορές τον μήνα
- ☐ Σπάνια

2) Τι ποσό ξοδεύετε κατά μέσο όρο σε κάθε σας επίσκεψη στο super market
(σημειώστε την απάντησή σας αριθμητικά):

3) Ποιά είναι τα κριτήρια με τα οποία αποφασίζετε σε ποιό σούπερ μάρκετ θα πάτε
(ανοιχτού τύπου) (π.χ. τιμές προϊόντων, απόσταση, προσφορές, ποιότητα
προϊόντων, ποικιλία προϊόντων, καθαριότητα σουπερ μάρκετ κλπ)
.....

4) Διαθέτετε κάρτα πόντων από κάποιο ή κάποια σούπερ μαρκετς; Αν ναι, από
ποιο/ά; (ανοιχτού τύπου)

- 5) Θα σας παρακαλούσαμε να σημειώσετε ποια σούπερ μαρκετ προτιμάτε συνήθως για τις αγορές των παρακάτω κατηγοριών προϊόντων. (Μπορείτε να επιλέξετε όσα σούπερ μαρκετ επιθυμείτε σημειώνοντας ένα X)

Κατηγορίες Προϊόντων	AB Βασιλόπουλος	LIDL	MART	Σκλαβενίτης	SYNKA	Χαλκιαδάκης
Φρούτα και λαχανικά						
Γαλακτομικά και Είδη ψυγείου						
Τυροκομικά						
Φρέσκο Κρέας						
Φρέσκο Ψάρι						
Αλλαντικά και Delicatessen						
Αρτοζαχαροπλαστείο						
Εποχιακά είδη						
Κατεψυγμένα τρόφιμα						
Ποτά, χυμοί, αναψυκτικά						
Συσκευασμένα τρόφιμα						
Υγεία και ομορφιά						
Οικιακή φροντίδα (καθαριστικά – χαρτικά)						
Σπίτι και κουζίνα						
Πρωινό, ροφήματα και σνακ						
Κατοικίδια						
Βρεφικά προϊόντα						
Αντισηπτική προστασία						
Βιολογικά – Vegan προϊόντα						
Κρητικά προϊόντα						

Μπορείτε να μας αναφέρετε αν εκτός από τα παραπάνω SM, αγοράζετε προϊόντα από άλλα καταστήματα; Ποια και γιατί (ανοικτή ερώτηση);

.....

Γ. Προτίμηση Super market (Επιλέξτε το κελί που αντιστοιχεί στην απάντησή σας)

1) Ποια είναι η γνώμη σας για τα σούπερ μαρκετ των παρακάτω αλυσίδων στα Χανιά;

Εταιρεία	Πολύ Αρνητική	Αρνητική	Μέτρια	Θετική	Πολύ Θετική
AB Βασιλόπουλος					
LIDL					
MART					
Σκλαβενίτης					
SYNKA					
Χαλκιαδάκης					

Δ. Εκτίμηση χαρακτηριστικών επιλογής super market

1) Σημειώστε, στο αντίστοιχο τετράγωνο, το βαθμό σημαντικότητας που αποδίδεται σε κάθε ένα από τα παρακάτω κριτήρια-χαρακτηριστικά, κατά τη διαδικασία επιλογής super market, χρησιμοποιώντας την κλίμακα 0-10 (με 0 σημειώνεται όταν δεν το λαμβάνεται καθόλου υπόψη ενώ με 10 αυτό με την μέγιστη σημαντικότητα)

Κριτήρια - χαρακτηριστικά	0	1	2	3	4	5	6	7	8	9	10
Τιμές προϊόντων											
Προσφορές											
Καθαριότητα και άνεση											
Ποιότητα προϊόντων											
Ποικιλία προϊόντων											
Απόσταση από το σπίτι											
Δυνατότητα διανομής προϊόντων στο σπίτι											
Εξυπηρέτηση από το προσωπικό											
Ασφάλεια στις συναλλαγές											
Ύπαρξη εξειδικευμένων προϊόντων (π.χ. τεχνολογικά προϊόντα, είδη ένδυσης κτλ)											
Ύπαρξη επαρκούς ποσότητας εμπορευμάτων											
Ύπαρξη ποικιλίας βιολογικών προϊόντων											
Ύπαρξη προϊόντων που μόλις έχουν κυκλοφορήσει											
Ύπαρξη χώρου στάθμευσης											
Ωράριο καταστήματος											
Ύπαρξη πολλών ταμείων σε λειτουργία											
Διαφήμιση											
Φήμη καταστήματος											
Ύπαρξη τοπικών προϊόντων											
Αν διαθέτετε κάρτα μόνους του super market											
Δυνατότητα αλλαγής ή επιδιόρθωσης ελαττωματικών προϊόντων											

- 2) Παρακαλούμε, να εκφράσετε τις προτιμήσεις σας για τις ακόλουθες εναλλακτικές επιλογές σούπερ μάρκετς (AB, LIDL, Mart, Σκλαβενίτης, SYNKA, Χαλκιαδάκης) με βάση τα παρακάτω χαρακτηριστικά-κριτήρια. Για την αξιολόγηση, χρησιμοποιήστε την ακόλουθη κλίμακα προτίμησης:

Πολύ κακή	1
Κακή	2
Μέτρια	3
Καλή	4
Πολύ καλή	5

Κριτήρια	AB Βασιλόπουλος	LIDL	MART	Σκλαβενίτης	SYNKA	Χαλκιαδάκης
Τιμές προϊόντων						
Προσφορές						
Καθαριότητα και άνεση						
Ποικιλία προϊόντων						
Ποιότητα προϊόντων						
Απόσταση από το σπίτι						
Δυνατότητα διανομής προϊόντων στο σπίτι						
Εξυπηρέτηση από το προσωπικό						
Ασφάλεια στις συναλλαγές						
Ύπαρξη εξειδικευμένων προϊόντων (π.χ. τεχνολογικά προϊόντα, είδη ένδυσης κτλ)						
Ύπαρξη επαρκούς ποσότητας εμπορευμάτων						
Ύπαρξη προϊόντων που μόλις έχουν κυκλοφορήσει						

Υπαρξη χώρου στάθμευσης						
Ωράριο καταστήματος						
Υπαρξη πολλών ταμείων σε λειτουργία						
Διαφήμιση						
Φήμη καταστήματος						
Υπαρξη τοπικών προϊόντων						
Ποικιλία βιολογικών προϊόντων						
Αν διαθέτετε κάρτα μπόνους του super market						
Δυνατότητα αλλαγής ή επιδιόρθωσης ελαττωματικών προϊόντων						

Λαμβάνοντας υπόψη σας όλες τις παραπάνω εκτιμήσεις σας στο σύνολο των κριτηρίων, σας παρακαλούμε να κατατάξετε τα super markets, με βάση τη σειρά προτίμησης για τις αγορές σας (Χρησιμοποιείστε την κλίμακα: 1: το πλέον προτιμώμενο, 2: το αμέσως επόμενο σε σειρά προτίμησης, κοκ. Σε περίπτωση που έχετε την ίδια προτίμηση μπορείτε να την εκφράσετε με απόδοση της ίδιας σειράς πχ. 2 και 2):

AB Βασιλόπουλος	LIDL	MART	Σκλαβενίτης	SYNKA	Χαλκιαδάκης

Δ. Παρατηρήσεις / σχόλια / προτάσεις που θέλετε να προσθέσετε για τα Super Markets της περιοχής σας (ανοικτή ερώτηση):

.....

.....

.....

.....

.....