



ATHENS UNIVERSITY
OF ECONOMICS
AND BUSINESS



ΜΕΤΑΠΤΥΧΙΑΚΟ
ΔΙΟΙΚΗΤΙΚΗ ΕΠΙΣΤΗΜΗ
& ΤΕΧΝΟΛΟΓΙΑ
MSc IN
MANAGEMENT SCIENCE
& TECHNOLOGY

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ
ΤΜΗΜΑ ΔΙΟΙΚΗΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

Μπάρκη Χαρίλαου

Ανάλυση χώρο-χρονικών δεδομένων κίνησης καταναλωτών σε κατάστημα
λιανεμπορίου

Επιβλέπων : Κατερίνα Πραματάρη Αναπληρώτρια Καθηγήτρια

Υποβληθείσα ως μέρος των απαιτήσεων για την απόκτηση
Μεταπτυχιακού Διπλώματος (MSc) στη Διοικητική Επιστήμη και Τεχνολογία

Αθήνα, Ιούλιος 2020



Η σελίδα αυτή είναι σκόπιμα λευκή.



Βεβαίωση εκπόνησης Διπλωματικής εργασίας

«Δηλώνω υπεύθυνα ότι η συγκεκριμένη μεταπτυχιακή εργασία για τη λήψη του μεταπτυχιακού τίτλου σπουδών του ΠΜΣ στη Διοικητική Επιστήμη και Τεχνολογία του Τμήματος Διοικητικής Επιστήμης και Τεχνολογίας του Οικονομικού Πανεπιστημίου Αθηνών έχει συγγραφεί από εμένα προσωπικά και δεν έχει υποβληθεί ούτε έχει εγκριθεί στο πλαίσιο κάποιου άλλου μεταπτυχιακού ή προπτυχιακού τίτλου σπουδών στην Ελλάδα ή το εξωτερικό. Η εργασία αυτή έχοντας εκπονηθεί από εμένα, αντιπροσωπεύει τις προσωπικές μου απόψεις επί του θέματος. Οι πηγές στις οποίες ανέτρεξα για την εκπόνηση της συγκεκριμένης διπλωματικής αναφέρονται στο σύνολό τους, δίνοντας πλήρεις αναφορές στους συγγραφείς, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο».

(Υπογραφή)

.....
Μπάρκης Χαρίλαος

Φοιτητής MSc στη Διοικητική Επιστήμη και Τεχνολογία



Περίληψη

Ο σκοπός της παρούσας διπλωματικής εργασίας είναι να δείξει πως μπορούν να αξιοποιηθούν νέες τεχνολογίες από συσκευές Διαδικτύου των Πραγμάτων (IoT) για να βοηθήσουν τους λιανέμπορους να κατανοήσουν τη συμπεριφορά του καταναλωτή μέσα στο κατάστημα. Στην εργασία παρατίθεται μια σειρά από πρακτικές εφαρμογές με βάση την ανάλυση θέσης (indoor location systems ILS). Στην συνέχεια παρουσιάζονται οι δυνατότητες που προσφέρει η ανάλυση και η εξόρυξη χωροχρονικών δεδομένων, προκειμένου να μπορέσουν οι λιανέμποροι να αναγνωρίσουν και να εντοπίσουν τις διαφορετικές ομάδες των καταναλωτών και τα χαρακτηριστικά αυτών. Με αυτόν τον τρόπο οι λιανέμποροι θα είναι σε θέση να ανέξουν την εμπειρία των καταναλωτών μέσα στο κατάστημα, να παρέχουν καινοτόμες υπηρεσίες στους πελάτες τους και ανταγωνιστικά προϊόντα που να ανταποκρίνονται καλύτερα στις ανάγκες των τελευταίων.

Πιο συγκεκριμένα για το σκοπό της ανάλυσης, στην εργασία παρουσιάζεται η μεθοδολογία Crisp-DM που εφαρμόστηκε σε δεδομένα συλλεγμένα μέσω ενός συστήματος εύρεσης θέσης από ένα κατάστημα λιανικής. Πραγματοποιήθηκε διερεύνηση των δεδομένων (περιγραφική στατιστική) και ακολούθως κατάτμηση των επισκέψεων των πελατών (clustering) για να εξαχθούν μοτίβα της αγοραστικής τους συμπεριφοράς. Επίσης χρησιμοποιήθηκαν Δείκτες Απόδοσης (KPIs) για να μετρήσουν και να οπτικοποιήσουν την επισκεψιμότητα του καταστήματος. Όλα τα ανωτέρω πραγματοποιήθηκαν προκειμένου να παρουσιαστεί η εφαρμογή της μεθόδου στην πράξη και να διερευνηθεί η συμβολή της στη λιανική αγορά.

Λέξεις Κλειδιά: Διαδίκτυο των Πραγμάτων, Λιανεμπόριο, Κατάτμηση Πελατών, Δείκτες Απόδοσης



Η σελίδα αυτή είναι σκόπιμα λευκή.



Abstract

The purpose of this dissertation is to show how modern technologies from Internet of Things (IoT) devices can be utilized to facilitate retailers understand in-store consumer behavior. This thesis presents a series of practical applications based on in-store analytics (indoor location systems ILS). The following are the capabilities offered by the analysis and processing of spatio-temporal data, so that retailers can identify the different groups of consumers and their characteristics. In this way retailers will be able to enhance the in-store consumer experience by providing innovative services to their customers and competitive products that best meet their needs.

More specifically for the purpose of the analysis, the thesis adopts the Crisp-DM methodology applied to data collected through a positioning system from a retail store. The data were analyzed (descriptive statistics) and then the customer visits were clustered to extract insights on shopper behavior. Key Performance Indicators (KPIs) were also used to measure and visualize store's traffic. All of the above were performed in order to present the application of the method in practice and to investigate its contribution to the retail industry.

Keywords: BLE Beacons, IoT, retail, KPIs, Customer Segmentation



Η σελίδα αυτή είναι σκόπιμα λευκή.



Πίνακας περιεχομένων

1	Εισαγωγή	1
1.1	Ανάλυση Δεδομένων Κίνησης Πελάτων Σε Κατάστημα (In-Store Customer Behavior)	1
1.2	Αντικείμενο Διπλωματικής.....	3
1.2.1	Συνεισφορά	4
1.3	Οργάνωση Κειμένου.....	4
2	Ερευνητικό Υπόβαθρο.....	6
2.1	Σχετικές Εργασίες Στη Χωρική Αναλυτική.....	7
2.2	Διάφορες Τεχνολογίες Εντοπισμού Θέσης.....	11
2.2.1	Τεχνολογίες Εύρεσης Θέσης	12
3	Μεθοδολογία	16
3.1	CRISP-DM Framework	17
3.2	Ανάλυση Συστάδων (Cluster Analysis).....	23
3.2.1	Μέτρα Απόστασης Data Preparation – Format Data -	24
3.2.2	Αλγόριθμοι Modeling	25
3.2.3	Επιλογή αριθμού συστάδων πελατών Build Model.....	28
4	Case Study	30
4.1	Business understanding Phase in Crisp-DM.....	31
4.2	Δεδομένα - Data Understanding Phase in Crisp-DM	33
4.2.1	Περιγραφή δεδομένων – Describe data Step	33
4.3	Προετοιμασία Δεδομένων -Data Preparation Phase in Crisp-DM	35
4.3.1	Δημιουργία νέων πεδίων δεδομένων – Construct data Step	35
4.3.2	Συγκεντρωτικά Χαρακτηριστικά -Dataset Description Step.....	36
4.3.3	Επιλογή Δεδομένων - Select Data Step	41
4.3.4	Οπτικοποίηση Των Μετακινήσεων Των Πελατών Μέσω Γράφων Explore Data Step	43
4.4	Μοντελοποίηση Modeling Phase in Crisp-DM	47
4.4.1	Τροποποίηση Δεδομένων Πριν Την Εισαγωγή Τους Στο Μοντέλο Construct Data And Transform Data Step	47
4.4.2	Σύγκριση Αλγορίθμων Για Την Ανάλυση Των Δεδομένων Assess Mode Step ..	48
4.4.3	Κατάτμηση Επισκέψεων (Visits Segmentation) Evaluation Step	50



5	Managerial implications.....	60
5.1	Key Performance Indicators (KPIs).....	61
5.1.1	Ορισμός Και Αξιοποίηση Των Key Performance Indicators (Kpis)	62
5.1.2	Οπτικοποιήση Των Key Performance Indicators (KPIs).....	65
5.2	Προσφερόμενες Δυνατότητες Της Τεχνολογίας BLE Beacons Στο Λιανικό Εμπόριο	74
5.2.1	Παραδείγματα Mobile Marketing Με Τη Χρήση Της Τεχνολογίας BLE Beacons Στο Λιανικό Εμπόριο	77
	Λοιποί Χώροι Εφαρμογής της Τεχνολογίας BLE Beacons.....	77
5.3	Αποδοχή Των Νέων Τεχνολογιών Από Τους Χρήστες	78
6	Επίλογος.....	79
6.1	Σύνοψη και Συμπεράσματα	79
6.2	Μελλοντικές Επεκτάσεις	80
	Βιβλιογραφία.....	82
Appendix 88		
	Appendix A Τεχνικές Λεπτομερείς	88
	Appendix B συστήματα εντοπισμού θέσης	89
	Appendix B.1 Τεχνικές εύρεσης θέσης	89
	Appendix B.2 Μετρικές σύγκρισης των συστημάτων εντοπισμού	90
	Appendix C.....	92



Ενρετήριο Εικόνων

Εικόνα 1: Πρόβλεψη της αγορά των IoT τεχνολογιών στο αμερικάνικο λιανεμπόριο (grandviewresearch.com) ...	11
Εικόνα 2: Παρουσιάζονται τα διαφορά στάδιά της μεθοδολογία CRISP-DM (Shearer, 2000)	17
Εικόνα 3: Μέθοδος επεξεργασία ακατέργαστων δεδομένων από την βάση δεδομένων (Kalgotra & Sharda, 2016) ...	20
Εικόνα 4: Συνοπτική παρουσίαση και εξήγηση των στόχων σε κάθε βήμα της μεθοδολογία (Chapman et al., 2000)	22
Εικόνα 5: Παράδειγμα δενδρογράμματος από (Tripathi et al., 2018)	27
Εικόνα 6: Παράδειγμα elbow οπού το βέλτιστο k=5 είναι το σημείο καμπής της γραφικής (Tripathi et al., 2018) ...	28
Εικόνα 7: Θέση των κατηγοριών στο κατάστημα	31
Εικόνα 8: Επισκόπηση συστήματος ευρέσεως θέσης με βάση την τεχνολογία BLE	33
Εικόνα 9: Κάτοψη του καταστήματος	33
Εικόνα 10: Διάγραμμα οντοτήτων-συσχετίσεων (ERD).....	35
Εικόνα 11: Αριθμός επισκέψεων ανά Μηνά	37
Εικόνα 12: Αριθμός επισκέψεων ανά Ημέρα	37
Εικόνα 13: Θηκόγραμμα των επισκέψεων ανά ημέρα της εβδομάδας.....	38
Εικόνα 14: Κατανομή της χρονικής διάρκειας όλων των επισκέψεων σε όλες τις κατηγορίες.....	38
Εικόνα 15: violin plot των χρόνων επίσκεψης ανά κατηγορία.....	40
Εικόνα 16: Συσχέτιση μεταξύ των κατηγοριών με μετρική τον χρόνο των πελατών ανά κατηγορία	41
Εικόνα 17:Sankey chart με τις μετακινήσεις των πελάτων στο κατάστημα.....	43
Εικόνα 18: Γράφος με τις μετακινήσεις των καταναλωτών από κατηγορία σε κατηγορία.	44
Εικόνα 19: Αριστερά (πίνακας) οι κατηγορίες και ο βαθμός των ακμών που τους αντιστοιχούν. Δεξιά ο Γράφος που αναπαριστά τις μετακινήσεις ανάμεσα στις κατηγορίες εκφρασμένες σε ποσοστά (εικόνα)	45
Εικόνα 20: Αριθμός συνδέσεων ανά κατηγορίαν	46
Εικόνα 21: Elbow plot για K_Medoids (αριστερά) και K_Modes (δεξιά)	48
Εικόνα 22: Αποτελέσματα της Silhouette score για K_medoids.....	49
Εικόνα 23: Αποτελέσματα της Silhouette Score για τον K modes.....	49
Εικόνα 24: Αριθμός επισκέψεων σε κάθε συστάδα	50
Εικόνα 25: Περιγραφικά στατιστικά για την πρώτη ομάδα επισκέψεων	51
Εικόνα 26: Περιγραφικά στατιστικά για τη δεύτερη ομάδα επισκέψεων.....	52
Εικόνα 27: Περιγραφικά στατιστικά για την τρίτη ομάδα επισκέψεων	53
Εικόνα 28: Περιγραφικά στατιστικά για την τέταρτη ομάδα επισκέψεων	54
Εικόνα 29: Περιγραφικά στατιστικά για την πέμπτη ομάδα επισκέψεων	55
Εικόνα 30: Περιγραφικά στατιστικά για την έκτη ομάδα επισκέψεων	56
Εικόνα 31: Περιγραφικά στατιστικά για την έβδομη ομάδα επισκέψεων	57
Εικόνα 32: Συγκεντρωτικά τα ποσοστά επίσκεψης ανά κατηγορία και συστάδα	59
Εικόνα 33: αριθμός των επισκέψεων των πελάτων.....	65
Εικόνα 34: Γράφημα χρονοδιαγραμμάτων των επισκέψεων σε επίπεδο ημέρας	66
Εικόνα 35: ιστόγραμμά με τις διάφορες τιμές της ώρα προσέλευσης στο κατάστημα	66
Εικόνα 36: Χρονική διάρκεια των επισκέψεων των πελατών ανά κατηγορία εκφρασμένη σε ώρες	67
Εικόνα 37: Ολικός αριθμός επισκέψεων ανά κατηγορία από τους πελάτες.....	68
Εικόνα 38: Επισκεψιμότητα ανά κατηγορία	69



Εικόνα 39: Μέσος αριθμός επανα-επισκέψεων ανά κατηγορία σε μια επίσκεψη	70
Εικόνα 40: Μέση διάρκεια επίσκεψης των πελατών ανά κατηγορία εκφρασμένη σε λεπτά	71
Εικόνα 41: Μέσο ποσοστό του χρόνου που δαπάνησαν οι πελάτες ανά επίσκεψη στην αντίστοιχη κατηγορία.....	72
Εικόνα 42: Σύγκριση των περιοχών μέσω του δείκτη GRP	73
Εικόνα 44: Εφαρμογές των Beacons πηγή (Minds.com)	74

Ευρετήριο Πινάκων

Πίνακας 1: Σχετικές εργασίες με της αντίστοιχες μεθόδους ανάλυσης σε χρονολογική σειρά.....	10
Πίνακας 2: Δεδομένα καταγραφής κίνησης πελατών στο κατάστημα	34
Πίνακας 3: Αντιστοιχία περιοχών καταστήματος με το αναγνωριστικό κωδικό της	34
Πίνακας 4: Αναθεωρημένος πίνακας με τα νέα πεδία.....	36
Πίνακας 5: Πίνακας με στατιστικά των πεδίων	36
Πίνακας 6: Διαγράμματα διασποράς και ευθείας ελαχίστων τετραγώνων για το χρόνο επίσκεψης και τον αριθμό των περιοχών επίσκεψης και μοναδικών περιοχών επίσκεψης	42
Πίνακας 7: Συγκεντρωτικά στατιστικά των ομάδων επισκέψεων.....	50
Πίνακας 8:Αντιστοιχία μετρικών σε ηλεκτρονικό κατάστημα με αυτών σε ένα φυσικό κατάστημα	64
Πίνακας 9:Στοχαστικός πίνακας των ποσοστών μετακίνησης/μετάβασης από μια (αρχική κατάσταση) κατηγορία (γραμμή) σε άλλη(τελική κατάσταση) κατηγορία (στήλη)	92



1

Εισαγωγή

1.1 Ανάλυση Δεδομένων Κίνησης Πελάτων Σε Κατάστημα

(In-Store Customer Behavior)

Τα Μεγάλα Δεδομένα (Big Data) και η Επιχειρηματική Αναλυτική (Business Analytics), αποτελούν πλέον αναπόσπαστο κομμάτι των διαδικασιών και της λήψης αποφάσεων των περισσότερων εταιρειών. Συγκεκριμένα έχουν δημιουργήσει νέες δυνατότητες και γνώση για τους πελάτες τους, που προηγουμένως δεν είχαν. Αυτές οι δυνατότητες και η γνώση πηγάζουν από τις νέες πήγες δεδομένων (π.χ. Web Analytics), τη συγκέντρωση και τη στατιστική ανάλυση αυτών των δεδομένων, μαζί με τη γνώση των αναλυτών/μάνατζερς της αγοράς, για την εκάστοτε επιχείρηση. Καθώς η συμπεριφορά των καταναλωτών αλλάζει, επισκεπτόμενοι ένα κατάστημα κάθε φορά για να εκπληρώσουν και μια διαφορετική τους ανάγκη, οι πελάτες γίνονται όλο και πιο απαιτητικοί. Γι' αυτό το λόγο, στο λιανεμπόριο βρίσκουν εφαρμογή αρκετές τεχνολογίες βασισμένες στην Εξόρυξη Γνώσης (Data Mining) και στην Επιχειρηματική Ευφυΐα (Business Intelligence ή BI), οι οποίες υποβοηθούν τις επιχειρήσεις στη λήψη αποφάσεων. Στόχος τους είναι να ανακαλύψουν τις ανάγκες του πελάτη τους ή και να τις προβλέψουν, καθώς έχουν αναγνωρίσει ότι η ικανοποίηση του πελάτη συνδέεται



άρρηκτα με την κερδοφορία τους. Πλέον, η συλλογή και αποθήκευση όσο το δυνατόν περισσότερων και διαφορετικών ειδών δεδομένων σχετικά με τους πελάτες τους, με σκοπό την ανάλυσή τους, αποτελεί για τις επιχειρήσεις καθημερινή πρακτική εδώ και χρόνια. Μέσα σ' αυτό το πλαίσιο, μια από τις μεγαλύτερες φιλοδοξίες τους είναι η εύρεση καινοτόμων τρόπων αξιοποίησης των συλλεγόμενων δεδομένων (Griva et al., 2018).

Η συλλογή και ανάλυση δεδομένων των πελατών, μέσω της ιστοσελίδας του καταστήματος (Web Analytics μέσω πλατφορμών όπως google analytics και adobe analytics), είναι πια συνηθισμένη στα διαδικτυακά καταστήματα λιανικής, όπου συλλέγονται δεδομένα για τις συναλλαγές που πραγματοποιήθηκαν, αλλά και για τις περιπτώσεις που ο καταναλωτής δεν προέβη σε αγορά (abandonment rate), που είναι και αυτές εξίσου σημαντικές προς ανάλυση, για την κατανόηση της συμπεριφοράς των καταναλωτών. Σκοπός της συλλογής και ανάλυσης δεδομένων είναι η περαιτέρω ανάπτυξη και βελτίωση της ιστοσελίδας, η βελτιστοποίηση της σελίδας προορισμού (landing page) και, εν τέλει, η αύξηση των μετατροπών (conversion rate). Αντίθετα, στα φυσικά καταστήματα, οι λιανέμποροι συνεχίζουν να επικεντρώνονται στην ανάλυση δεδομένων των συναλλαγών (δεδομένα από τα τερματικά POS του καταστήματος). Ενώ έχουν αναπτυχθεί και αναλυθεί στρατηγικές για την ανάλυση των επισκέψεων των πελάτων στα καταστήματα για δεκαετίες (π.χ. μέσω της καταμέτρησης, της παρακολούθησης από αναλυτές, δείγματος πελατών (focus groups), χρήσης ερωτηματολογίων), αυτές οι τεχνικές είναι μη (ή εν μέρη) αυτοματοποιημένες, χρονοβόρες και δαπανηρές στη συλλογή των δεδομένων (Liebig & Wagoum, 2012).

Η τέταρτη βιομηχανική επανάσταση, όπως αναφέρεται τα τελευταία χρόνια η εξάπλωση των συσκευών Διαδικτύου των Πραγμάτων (Internet of Things) από τη βιομηχανία μέχρι και την καθημερινή ζωή των ανθρώπων, προσφέρεται για τη συλλογή αυτού του τύπου δεδομένων, κάτι που τα προηγούμενα χρόνια δεν ήταν εύκολα υλοποιήσιμο.

Το κενό έρχεται να συμπληρώσει και η Χωρική Αναλυτική (Location Analytics), η οποία αναφέρεται στην ανάλυση χωρικών και χρονικών δεδομένων (από εδώ και στο εξής χωροχρονικών), προκειμένου να ανευρεθούν σχέσεις (insights) και μοτίβα (patterns) που ερμηνεύονται τη συμπεριφορά του πελάτη, αναλύοντας την κίνησή του μέσα στο κατάστημα (Senior et al., 2007).

Η παρακολούθηση της κίνησης των καταναλωτών μέσα στο κατάστημα, παρέχει πληροφορίες σε συγκεντρωτικό επίπεδο για τις περιοχές που επισκέπτονται πιο συχνά, που περνάνε το πιο πολύ χρόνο και τις περιοχές με την μικρότερη επισκεψιμότητα. Οι πληροφορίες αυτές μπορούν να αξιοποιηθούν για την εκτίμηση και παρακολούθηση δεικτών απόδοσης (KPIs), προκειμένου οι λιανέμποροι να οργανώσουν καλύτερα τις προωθητικές τους ενέργειες και τη λειτουργία του καταστήματος (Oosterlinck et al., 2017).

Επιπλέον, η τεχνολογία των IoTs, δίνει τη δυνατότητα στους λιανέμπορους να συγκεντρώσουν εξατομικευμένα δεδομένα για τον κάθε πελάτη και να τα συνδυάσουν με τα δεδομένα αγορών (δεδομένα από τα POS μέσω ενός προγράμματος ανταμοιβής).

Οι δυνατότητες αξιοποίησης των συνδυασμών των δεδομένων είναι πάρα πολλές. Μπορούν να παρέχουν ένα εξατομικευμένο σύστημα σύστασης προϊόντων, όπως παρέχεται ήδη στα ηλεκτρονικά καταστήματα, με σκοπό να βελτιώσουν την εμπειρία των πελατών μέσα στο κατάστημα και να αυξήσουν τις πωλήσεις τους. Δηλαδή να χρησιμοποιήσουν τεχνικές που εφαρμόζονται στα ηλεκτρονικά καταστήματα, ώστε να φέρουν την εμπειρία του ηλεκτρονικού καταστήματος μέσα στο φυσικό και να τα «ενοποιήσουν» κατά κάποιο τρόπο, ώστε να δημιουργήσουν ένα πανκαναλικό (omnichannel) αγοραστικό περιβάλλον, σε πραγματικό χρόνο, μέσα στο κατάστημα (Yaeli et al., 2014). Αυτό, τους δίνει τη δυνατότητα να προχωρήσουν και σ' ένα πρόγραμμα εξατομικευμένης τιμολογιακής διάρθρωσης, με το οποίο μπορούν να προσφέρουν εξατομικευμένα κουπόνια στον κάθε πελάτη, για τα προϊόντα που είναι πιο πιθανό να αντιδράσει αυτός θετικά, ώστε να αυξήσουν τις αυθόρυμητες αγορές και τις παράλληλες πωλήσεις (cross sale).

Απ' όλα τα παραπάνω, γίνεται κατανοητό ότι η ανάλυση των δεδομένων των επισκέψεων των πελατών σ' ένα φυσικό κατάστημα, μπορεί να προσφέρει παρόμοια αξία για την κατανόηση και βελτίωση των υπηρεσιών και των ποσοστών μετατροπής για το λιανικό εμπόριο, όπως έχει κάνει το web-analytics στα ηλεκτρονικά καταστήματα.

1.2 Αντικείμενο Διπλωματικής

Ο στόχος της παρούσας εργασίας είναι, κατ' αρχάς, η απαρίθμηση των τεχνολογιών για τον εντοπισμό θέσης σε εσωτερικούς χώρους και η γνωριμία του αναγνώστη με αυτές. Αναλύονται τα χαρακτηριστικά κάθε τεχνολογίας και γίνονται οι αναγκαίες συγκρίσεις μεταξύ τους, καθώς και οι πρακτικές εφαρμογές αυτών. Έπειτα η εργασία στοχεύει στην ανάλυση αυτού του τύπου χωροχρονικών δεδομένων, και πως μπορούν να χρησιμοποιηθούν τα αποτελέσματα της ανάλυσης τέτοιου τύπου δεδομένων, ώστε να παράγουν αξία για ένα κατάστημα, δηλαδή τι μετρικές μπορούν να δημιουργηθούν και πως μπορούν να οπτικοποιηθούν τα αποτελέσματα. Αυτό πραγματοποιείται μέσω βιβλιογραφικών αναφορών και μέσω της ανάλυσης δεδομένων από ένα ιδεατό κατάστημα.



1.2.1 Συνεισφορά

Για το σκοπό αυτό, στόχος της παρούσας διατριβής είναι να απαντήσει στα ακόλουθα ερευνητικά ερωτήματα:

- Πώς μπορεί ένας οργανισμός να αξιοποιήσει τα χωροχρονικά δεδομένα που συλλέγει με τεχνολογίες ανάλυσης θέσης, για να υποστηρίξει τη λήψη αποφάσεων του (decision making process); Με άλλα λόγια, πώς μπορούν να χρησιμοποιηθούν αυτά τα δεδομένα για την ανάλυση της πελατειακής βάσης; Πιο συγκεκριμένα: «Ποιές είναι οι περιοχές που οι καταναλωτές προτιμούν στο συγκεκριμένο κατάστημα;»
- Ποιες μετρικές μπορούν να οριστούν για την ανάλυση της θέσης των πελάτων στο φυσικό κατάστημα; Πώς μπορούν αυτές οι μετρικές να αξιοποιηθούν για να μελετηθεί η συμπεριφορά κίνησης των πελατών και να συνεισφέρουν στην καλύτερη κατανόηση τους;

Σημείωση: Με τον όρο «περιοχές», εννοούμε κατηγορίες προϊόντων, όπως αυτές έχουν τοποθετηθεί μέσα στο κατάστημα από τον λιανέμπορο. Συνεπώς μέσα στην παρούσα εργασία, εναλλάσσεται η χρήση των όρων «περιοχές» και «κατηγορίες προϊόντων», αλλά ουσιαστικά σημαίνουν το ίδιο, καθώς οι «περιοχές» αντικατοπτρίζουν τη χωρική υπόσταση των προϊόντων στο κατάστημα.

1.3 Οργάνωση Κειμένου

Τα κεφάλαια της διπλωματικής οργανώνονται ως εξής:

Στο πρώτο κεφάλαιο, της διπλωματικής εργασίας αναφέρονται περιληπτικά ο σκοπός και οι στόχοι της παρούσας διπλωματικής.

Στο δεύτερο κεφάλαιο, γίνεται αναφορά σε σχετικές εργασίες σχετικά με τους τρόπους ανάλυσης (location analysis) αυτών των χωρικών και χρονικών δεδομένων, τα οποία συλλέγονται από συστήματα εντοπισμού θέσης. Ακολουθεί και μια εισαγωγή στις τεχνολογίες εντοπισμού θέσης Indoor Positioning Systems (IPS).



Στο τρίτο κεφάλαιο, γίνεται αναφορά στη μεθοδολογία που εφαρμόζεται στην ανάλυση δεδομένων. Στην παρούσα διπλωματική, έχει επιλεχτεί ως μεθοδολογία η CRISP-DM (Shearer, 2000). Επιπλέον, αναπτύσσεται και το σχετικό θεωρητικό υπόβαθρο που θα χρειαστεί ο αναγνώστης για την ανάλυση που θα ακολουθήσει.

Στο τέταρτο κεφάλαιο, περιγράφονται τα δεδομένα κίνησης σε κατάστημα, τα οποία χρησιμοποιούνται στο κομμάτι της ανάλυσης της διπλωματικής εργασίας. Παρουσιάζεται η εφαρμογή της μεθοδολογίας CRISP-DM στα υπό μελέτη δεδομένα. Συγκεκριμένα αναφέρονται οι αναγκαίες μετατροπές στα δεδομένα και η περιγραφική τους ανάλυση, υλοποιείται ανάλυση συστάδων των επισκέψεων, ώστε να εντοπιστούν οι αγοραστικές αποστολές (shopping missions) και παραθέτονται τα αποτελέσματά τους καθώς και σχολιασμός αυτών.

Στο πέμπτο κεφαλαίο, γίνεται αναφορά σε μετρικές ή δείκτες (KPIs), που θα χρειαστούν για την ανάλυση τέτοιου τύπου δεδομένων. Στη συνέχεια, παρουσιάζονται τα συμπεράσματα που μπορούν να εξαχθούν από αυτούς τους δείκτες και γίνεται μια αναφορά της εφαρμογής των παραπάνω τεχνολογιών εντοπισμού θέσης, σε ορισμένες real world cases, όπως αυτές έχουν εφαρμοστεί στον κόσμο του λιανεμπορίου.

Επιπλέον, γίνεται και μια αναφορά – σχολιασμός στο πως θα μπορούσε να εφαρμοστεί ο τρόπος ανάλυσης που αναπτύχθηκε στο προηγούμενο κεφάλαιο, καθώς και στους σχετικούς περιορισμούς που θα μπορούσαν να υπάρξουν σχετικά με τη συλλογή και την ανάλυση των δεδομένων.

Στο έκτο κεφάλαιο, γίνεται μία σύνοψη της όλης εργασίας και συνοψίζονται τα σχετικά συμπεράσματα, καθώς και οι δυνατότητες για την περαιτέρω επέκταση της έρευνας πάνω στον συγκεκριμένο τομέα.

2

Ερευνητικό Υπόβαθρο

Σ' αυτό το κεφάλαιο θα παρουσιαστούν σχετικές εργασίες αναφορικά με τους τρόπους ανάλυσης (location analysis) των δεδομένων που συλλέγονται από τα ανωτέρω συστήματα IPS. Στη δεύτερη ενότητα του κεφαλαίου παρουσιάζονται οι διάφορες τεχνολογίες εντοπισμού θέσης (indoor positioning Systems IPS), θα γίνει ιδιαίτερη αναφορά στα BLE Beacons καθώς και στη σύγκρισή τους με άλλες τεχνολογίες

2.1 Σχετικές Εργασίες Στη Χωρική Αναλυτική

Για την ανάλυση των χωρικών δεδομένων δηλαδή της ανάλυσης της κίνησης των πελατών σ' ένα κατάστημα υπάρχει ένας ξεχωριστός τομέας στην επιχειρηματική αναλυτική που ονομάζεται χωρική αναλυτική (location analytics), με σκοπό να προσφέρει εξειδικευμένη στατιστική ανάλυση διαφόρων χρονικών και γεω/χωρικών δεδομένων σχετικά με την τοποθεσία του χρήστη (ArcNews, 2012). Τέτοια αναλυτικά στοιχεία μπορούν να παράγουν χρήσιμες πληροφορίες προκειμένου να ανευρεθούν σχέσεις (insights) και μοτίβα (patterns) που ερμηνεύονται τη συμπεριφορά των πελατών μέσα σε ένα κατάστημα στην προκείμενη περίπτωση. Οι Yaeli et al. τονίζουν τη δυνατότητα της χωρικής ανάλυσης στο να προσφέρει προστιθέμενη αξία στις επιχειρήσεις λιανικής και να υποστηρίξει τους managers ή/και marketers στη στρατηγική λήψης αποφάσεων. Επίσης, αναγνωρίζουν (Yaeli et al.) ότι παρέχει αξία και στους δύο συμμετέχοντες (stakeholders). Πρώτον, στο κατάστημα δίνει τη δυνατότητα να παρέχει μια πιο προσωποποιημένη εμπειρία αγοράς στους πελάτες του ώστε να ενδυναμώσει την πιστότητά τους και να αυξήσει τα έσοδά του. Δεύτερον, οι πελάτες που επισκέπτονται το φυσικό κατάστημα και χρησιμοποιούν τις υπηρεσίες της τεχνολογίας που έχει αναπτυχθεί για την εφαρμογή της χωρικής ανάλυσης, διευκολύνονται στις αγορές τους (Yaeli et al., 2014).

Η διαπίστωση ότι η διάταξη ενός καταστήματος επηρεάζει την αγοραστική συμπεριφορά των καταναλωτών έχει διατυπωθεί από τους ερευνητές από το 1966 (Oosterlinck et al., 2017). Από τους πρώτους που χρησιμοποίησαν σύγχρονα μέσα για να καταγράψουν την κίνηση των καταναλωτών μέσα σε κατάστημα ήταν ο Sorenson που έβαλε RFID tags σε καρότσια σουπερμάρκετ. Με τη συλλογή 200.000 δεδομένων επίσκεψης παρατήρησε ότι οι πελάτες που κινούνται στους διαδρόμους του σουπερμάρκετ με φορά αντίστροφη του ρολογιού είχαν καλάθια αξίας 2 δολαρίων παραπάνω κατά μέσο όρο. Επιπλέον, αν η είσοδος του καταστήματος ήταν στη δεξιά πλευρά ευνοούσε των ανωτέρω κίνηση (H Sorenson, 2003).

Πρωτοπόροι στην ανάλυση χωρικών δεδομένων ήταν και οι Larson et al. που χρησιμοποίησαν RFIDs τοποθετημένα σε καρότσια αμερικανικού σουπερμάρκετ και κατέγραψαν τις συντεταγμένες της θέσης τους. Το δείγμα τους αποτελούνταν από 27.000 καταγεγραμμένες επισκέψεις. Με τη χρήση του K-medoids ομαδοποίησαν τις χωρικές συντεταγμένες σε 14 διαφορετικά μονοπάτια. Πρώτο κριτήριο ομαδοποίησης των επισκέψεων σε διαφορετικά μονοπάτια ήταν ο χρόνος παραμονής των πελατών στο κατάστημα. Δεύτερο κριτήριο ομαδοποίησης ήταν η μεταβολή των χωρικών συντεταγμένων των καροτσιών. Συμπέραν ότι τα περισσότερα ξεχωριστά μονοπάτια οφείλονταν σε επισκέψεις σε διαφορετικούς διάδρομους και κατέρριψαν την ως τότε άποψη ότι οι πελάτες περιπλανιούνται σ' όλο το κατάστημα (Larson et al., 2005).



Παρόμοια έρευνα πραγματοποιήθηκε από τους Sano et al. σε σουπερμάρκετ στην Ιαπωνία. Χρησιμοποίησαν δεδομένα κίνησης για να εξάγουν 9 ξεχωριστά μονοπάτια. Στη συνέχεια τα ανέλυσαν, αξιοποιώντας τα δημογραφικά δεδομένα, τα δεδομένα πωλήσεων και το χρόνο παραμονής ανά περιοχή. (Sano, Tsutsui, Yada, & Suzuki, 2016)

Οι Oosterlinck et al. χρησιμοποιήσαν δέκτες Bluetooth για να παρακολουθήσουν τις κινήσεις των επισκεπτών στα καταστήματα ενός εμπορικού κέντρου. Συγκεκριμένα παρακολουθούσαν συσκευές κινητών τηλεφώνων που οι ιδιοκτήτες τους είχαν ανοικτές τις κεραίες Bluetooth. Όσον αφορά στη σειρά με την οποία επισκέπτονταν τα καταστήματα, παρατήρησαν ότι οι πελάτες είναι πολύ πιθανό να επιλέξουν ένα γειτονικό κατάστημα σ' αυτό που βρίσκονταν ήδη. Επίσης προχώρησαν σε ιεραρχική ομαδοποίηση για να βρουν το σκοπό των επισκέψεων. Βρήκαν 5 χαρακτηριστικές κατηγορίες πελατών, οι 3 κατηγορίες είχαν επισκεφτεί σχεδόν αποκλειστικά ένα συγκεκριμένο κατάστημα, ενώ η άλλη ομάδα είχε επισκεφτεί το εμπορικό κέντρο μόνο για φαγητό και μια γενική κατηγορία είχε επισκεφτεί μια πληθώρα καταστημάτων (Oosterlinck et al., 2017).

Οι Kaneko and Yada συνέλεξαν δεδομένα κίνησης πελατών μέσα σε κατάστημα supermarket στην Ιαπωνία με καρότσια με πομπούς RFIDs και δεδομένα από POS. Χρησιμοποίησαν τη μορφοκλασματική διάσταση (Fractal dimension, θεωρία από fractal) για να ποσοτικοποιήσουν την πολυπλοκότητα και την τυχαιότητα (πόσο πολύ περιηγήθηκαν στις περιοχές του καταστήματος) του μονοπατιού που διένυσαν οι πελάτες. Τη μετρική που προέκυψε από τη μελέτη, την συμπεριέλαβαν στο μοντέλο πωλήσεών τους. Αφού χώρισαν τους πελάτες σε δύο ομάδες, βάση της πολυπλοκότητας, κατέληξαν στο συμπέρασμα ότι η ομάδα των πελατών με το πιο πολύπλοκο μονοπάτι, πέρα από το ότι είχαν μεγαλύτερο χρόνο διαμονής και μεγαλύτερο αριθμό προϊόντων στα καλάθια τους είχαν και μεγαλύτερης αξίας καλάθια και ήταν παντρεμένες γυναίκες που ψώνιζαν φρέσκα υλικά για μαγείρεμα. Ενώ τα λιγότερο πολύπλοκα μονοπάτια αποτελούνταν από ανύπαντρους άντρες που ψώνιζαν έτοιμα γεύματα και προϊόντα. Επίσης το μοντέλο τους έδειξε ότι τα φρέσκα προϊόντα για προετοιμασία γεύματος σχετίζονται με την πολυπλοκότητα των μονοπατιών ενώ για έτοιμα προϊόντα έδειξε ότι η συσχέτισή τους ήταν μη στατιστικά σημαντική. (Kaneko & Yada, 2016).

Παρόμοια έρευνα πραγματοποίησαν και οι Fujino et al. που ήθελαν να μελετήσουν πως επηρεάζει το μήκος του μονοπατιού (εκφρασμένο σε μέτρα) του καταναλωτή, την αγοραστική του συμπεριφορά. Για αυτό το λόγο, εγκατέστησαν Rfids σε καρότσια σουπερμάρκετ. Με τα δεδομένα κίνησης που τα συνδύασαν με δεδομένα Pos (εγκατέστησαν rfids στα ταμεία και συγχρόνισαν τα καρότσια με τις αγορές) παρατήρησαν ισχυρή θετική συσχέτιση του μήκους του μονοπατιού με το χρόνο παραμονής στο κατάστημα αλλά ασθενή θετική με το μέγεθος του καλαθιού. Να σημειωθεί ότι το μικρό δείγμα της έρευνας επηρεάζει την αξιοπιστία του αποτελέσματος (Fujino et al., 2014).



Οι Hui et al. ήθελαν να μελετήσουν τα μονοπάτια κίνησης των πελατών σε σουπερμάρκετ. Χρησιμοποίησαν αλγόριθμο από το Πρόβλημα του Πλανόδιου Πωλητή (Travelling Salesman Problem) για να βρουν το βέλτιστο μονοπάτι βάση των προϊόντων του καλαθιού του πελάτη. Σύγκριναν τη βέλτιστη διαδρομή με τη διαδρομή που πραγματοποίησε ο πελάτης και βρήκαν απόκλιση, την οποία ερμήνευσαν ότι οφείλεται στη διαφορά στη σειρά αγοράς των προϊόντων και στην επιπλέον απόσταση. Ανάμεσα στα συμπεράσματα που έβγαλαν ήταν ότι οι διαδρομές που αποκλίνουν περισσότερο από τη βέλτιστη διαδρομή σχετίζονται με τα μεγαλύτερα καλάθια αγορών και ότι ένα μεγαλύτερο κομμάτι της διαδρομής του πελάτη μέσα στο κατάστημα δεν σχετίζεται τις αγορές του.(Hui, Fader, & Bradlow, 2009)

Οι Jung et al προτείνουν μια νέα μέθοδο ανάλυσης των μονοπατιών των καταναλωτών που περιλαμβάνει έναν αλγόριθμο ομαδοποίησης των μονοπατιών αγοράς που συγκρίνει την ακολουθία των μονοπατιών με τη μέθοδο LCSS (longest common subsequence) για να βρει το βαθμό ομοιότητας που έχουν μεταξύ τους τα μονοπάτια και να τα ομαδοποιήσει. Αυτή η μέθοδος που προτείνουν, παρέχει πληροφορίες απαραίτητες για τον προσδιορισμό της συχνότητας αγοράς των πελατών και τον καθορισμό των σημαντικών σημείων μέσα στο κατάστημα. Μέσα από αυτά τα συμπεράσματα θα είναι εφικτό να γίνει περισσότερο κατανοητή η καταναλωτική συμπεριφορά των καταναλωτών και να προσδιοριστούν οι νέες θέσεις των προϊόντων μέσα στο κατάστημα. Αυτή η μέθοδος χρησιμοποιήθηκε με πραγματικά δεδομένα από ένα σουπερμάρκετ στη Σεούλ και προσδιόρισαν με μεγάλη επιτυχία τα σημεία υψηλής και χαμηλής επισκεψιμότητας και το μοτίβο κίνησης των πελατών.(Jung, Kwon, & Lee, 2012)

Για την ανάλυση της απόδοσης των διάφορων προϊόντων μέσα στο κατάστημα οι Ferracuti et al. έλαβαν υπόψιν τους τις μέχρι τότε γνωστές μετρικές και εισήγαγαν μια νέα μετρική που σχετίζεται με το εμβαδό του κάθε τμήματος/περιοχής του καταστήματος. Ως εκ τούτου η διαίρεση των μετρικών μιας περιοχής με το εμβαδό της, κανονικοποιεί τις μετρικές και γίνεται καλύτερη απεικόνιση της απόδοσης των προϊόντων ανά περιοχή καταστήματος. Προχώρησαν στην κατασκευή ενός πιθανολογικού μοντέλου επισκεψιμότητας των μονοπατιών που τους επέτρεψε να ταξινομήσουν όλα τα πιθανά μονοπάτια. Αυτή η ταξινόμηση έχει πολλές πρακτικές εφαρμογές στο μάρκετινγκ και τη διαφήμιση (Ferracuti et al., 2019).

Ερευνητές	Τεχνολογία	Μετρήσιμες Μετρικές	Μέθοδοι Ανάλυσης
(Larson et al., 2005)	RFID	χρόνος και αριθμός πελάτων ανά περιοχή του καταστήματος	κατάτμησης των πελατών βάση των μονοπατιών που διάνυσαν
(Senior et al., 2007)	CCTV	χρόνος και αριθμός πελάτων ανά περιοχή του καταστήματος	Περιγραφική στατιστική, heat maps
(Herb Sorensen, 2009)	RFID	χρηματικό ποσό, χρόνος “Second per Dollar”	Correlation analysis
(Hui et al., 2009)	RFID	μήκους του μονοπατιού, το χρόνο παραμονής μέγεθος καλαθιού	Correlation analysis, TSP
(Jung et al., 2012)	RFID	μήκους του μονοπατιού, σειρά επίσκεψης των περιοχών	Sequence pattern matching (LCSS), sequence clustering
(Fujino et al., 2014)	RFID	μήκους του μονοπατιού, το χρόνο παραμονής μέγεθος καλαθιού	Περιγραφική στατιστική, συσχέτιση
(Yaeli et al., 2014)	WIFI/BLEs	χρόνος και αριθμός πελάτων ανα περιοχές και καταστήματα επίσκεψης, ώρα εισόδου εξόδου	Περιγραφική στατιστική, flow diagrams
(Kaneko & Yada, 2016)	RFID	χρόνος και απόσταση που διένυσαν στις περιοχές του καταστήματος οι πελάτες, δεδομένα πωλήσεων	Ανάλυση αγοραστικής συμπεριφοράς βάση της πολυπλοκότητας του μονοπατιού που διένυσαν Fractal Dimension
(Oosterlinck et al., 2017)	Bluetooth	Μη προσωποποιημένα δεδομένα, Περιοχές καταστημάτων, χρόνο ανά κατάστημα	Περιγραφική στατιστική, flow diagrams, Ιεραρχική ομαδοποίηση των επισκέψεων
(Ferracuti et al., 2019)	RFID	Περιοχές καταστήματος, ταχύτητα πελάτων, Εμβαδόν/χρόνο επίσκεψης	Περιγραφική στατιστική, πιθανολογικό μοντέλο για την ανεύρεση των συχνότερων διάδρομων

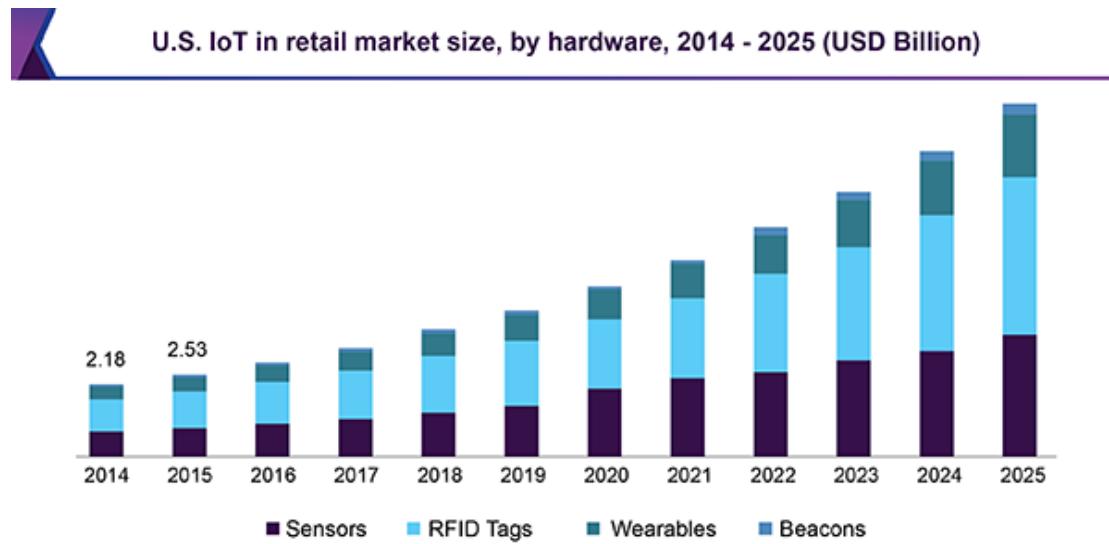
Πίνακας 1: Σχετικές εργασίες με της αντίστοιχες μεθόδους ανάλυσης σε χρονολογική σειρά

2.2 Διάφορες Τεχνολογίες Εντοπισμού Θέσης

Η έλευση του διαδικτύου των πράγματων (IoTs) και η έρευνα σε MEMS¹, κατάφεραν να ρίξουν το κόστος και το μέγεθος των αισθητήρων (sensors π.χ. επιταχυνσιόμετρα αισθητήρες υγρασίας, μαγνητικού πεδίου) και να βρουν καινοτόμες εφαρμογές για ευρεία εφαρμογή σε διάφορες βιομηχανίες και αγορές. Τα IoTs βρίσκουν εφαρμογές από μεγάλους οργανισμούς στην τήρηση της αποθήκης, στη μεταφορά των εμπορευμάτων, στην αυτοματοποίηση διεργασιών, στην παρακολούθηση των μηχανών παραγωγής και βελτιστοποίησης της παραγωγής μέχρι τις κινητές συσκευές και τις έξυπνες οικιακές συσκευές των σπιτιών του μέλλοντος.

Μετά τις βιομηχανίες ο πιο γοργά αναπτυσσόμενος τομέας εφαρμογής τους είναι το εμπόριο. Η παγκόσμια αγορά των IoT στην λιανική είχε υπολογιστεί στα 16,36 δις δολάρια, και προβλέπεται να παρουσιάσει σύνθετο ετήσιο ποσοστό αύξησης (CAGR) 21,5% δηλαδή να ξεπεράσει τα 90 δις δολάρια μέχρι το 2025 (grandviewresearch.com).

Αντίστοιχα στην αγορά της Αμερικής τα IoTs, μέχρι το 2025 προβλέπεται να αγγίξουν τα 40,5 δις δολάρια εκ των οποίων το μεγαλύτερο ποσοστό της λιανικής αναμένεται να το κατέχουν τα RFIDs με 11,6 δις (Statista.com).



Εικόνα 1: Πρόβλεψη της αγορά των IoT τεχνολογιών στο αμερικανικό λιανεμπόριο (grandviewresearch.com)

¹ Microelectromechanical systems

2.2.1 Τεχνολογίες Εύρεσης Θέσης

Πολλές επιστημονικές εργασίες έχουν γραφτεί για διαφορετικές τεχνολογίες σχετικά με τα IPS. Οι Liu και Brena στις έρευνες τους αναλύουν συγκεντρωτικά για διάφορες τεχνικές και προσεγγίσεις για το IPS. Παρουσιάζουν και συζητούν τεχνικές μέτρησης και τοποθεσίας όπως ο τριγωνισμός (triangulation), ο τριπλευρισμός (trilateration), fingerprinting και την εγγύτητα (proximity). Συζητούν επίσης έναν τρόπο αξιολόγησης του IPS ως προς την ακρίβεια, το κόστος και την ευρωστία μεταξύ άλλων. Παρουσιάζουν και συγκρίνουν μια ταξινόμηση διαφορετικών ασύρματων τεχνολογιών που μπορούν να χρησιμοποιηθούν για την κατασκευή IPS (Liu et al., 2007; Brena et al., 2017). Οι Seco et al. παρουσιάζουν αναλυτικά τις μαθηματικές προσεγγίσεις για συστήματα με βάση τις ραδιοσυχνότητες (RF), υπάρχουν τέσσερις κυρίες κατηγορίες προσεγγίσεων που μπορούν να χρησιμοποιηθούν και είναι:

Μέθοδοι βασισμένοι στην γεωμετρία όπως Time of Flight (ToF), τριγωνοποίηση, τριπλευρισμός. Μέθοδοι ελαχιστοποίησης μια συνάρτησης κόστους, fingerprinting και μεθόδου βασισμένες στην μπεϋζιανή στατιστική όπως Kalman filters (Seco, Jiménez, Prieto, Roa, & Koutsou, 2009).

Παρακάτω η εργασία αναφέρεται στις τεχνολογίες και τεχνικές που μπορούν να χρησιμοποιηθούν για τη δημιουργία ενός IPS. Αυτή η πληροφορία είναι πολύτιμη και χρειάζεται να γίνει η κατάλληλη έρευνα πριν την υλοποίηση οπουδήποτε συστήματος.

GPS

Η λειτουργία των συστημάτων εντοπισμού GPS στηρίζεται στη μέθοδο του τριπλευρισμού². Κάθε δορυφόρος μεταδίδει ένα ηλεκτρομαγνητικό κύμα το οποίο είναι μοναδικό για κάθε δορυφόρο. Για να εντοπίσει τη θέση του ένας δέκτης GPS, λαμβάνει σήματα από τρεις δορυφόρους και υπολογίζει την ακριβή απόστασή του από κάθε δορυφόρο. Η ακριβής τοποθεσία βρίσκεται από την τομή των τριών σφαιρών με απόκλιση λίγων μέτρων βάσει αυτών των αποστάσεων. Άλλα λόγω μεγάλων σφάλματων δεν ενδείκνυται για χρήση σε εσωτερικούς χώρους (Zeimpekis et al., 2003). Για το σκοπό αυτό, έχουν προταθεί άλλες λύσεις όπως το A-GPS(assisted) το οποίο χρησιμοποιεί και άλλα δεδομένα θέσης από άλλες τεχνολογίες ή η χρήση πολλαπλών εσωτερικών κεραιών αναμετάδοσης τους GPS (Rizos et al., 2010). Οι τεχνικές έχουν μεγάλο κόστος και χρειάζονται άλλες τεχνολογίες.

² Για καλύτερη απόδοση χρησιμοποιεί τον πολύ-πλεύρισμο δηλαδή 4 η και παραπάνω δορυφόρους για καλύτερη ακρίβεια

Wi-Fi

Το Wi-Fi ήταν μια αποδεδειγμένη λύση για τη σύνδεση πολλών υπολογιστών. Σε αντίθεση με το GPS (και τις κεραίες τηλεφωνίας που δουλεύουν με παρόμοιο τρόπο) το Wi-Fi μπορεί να βασιστεί στο χρόνο ή στη μέτρηση της απόστασης για να βρει τη θέση ενός αντικειμένου. Το Wi-Fi μπορεί να υπολογίσει, κατά προσέγγιση, τη θέση χρησιμοποιώντας τη λαμβανόμενη ισχύ σήματος - received signal strength RSS- από κάθε αντικείμενο. Και μπορεί να χρησιμοποιήσει την τεχνική fingerprinting μαζί με αλγορίθμους μηχανικής μάθησης για τη βελτίωση της ακρίβειας θέσης. Το πρωτόκολλο Wi-Fi κατασκευάστηκε για να συνδέονται υπολογιστές μέσω ασύρματων τοπικών δικτύων οπότε έχει και υψηλή ταχύτητα μετάδοσης και ισχυρό σήμα. Τα μειονεκτήματα του Wi-Fi είναι ότι είναι σχετικά ακριβό στη χρήση καθώς χρειάζονται πολλοί αναμεταδότες Access Points (AP) και έχει υψηλή ισχύ κατανάλωσης, κάτι που για φορητές εφαρμογές που λειτουργούν σε εικοσιτετράωρη βάση, κάνει το πρωτόκολλο Wi-Fi αναποτελεσματικό.

RFID

Τα RFIDs έχουν χρησιμοποιηθεί από πολλούς ερευνητές για να παρακολουθήσουν την κίνηση των πελατών, διακρίνονται δε σε δύο κατηγορίες, τα active και τα passive tags.

Τα active χρειάζονται μπαταρία καθώς χρησιμοποιούν ένα ολοκληρωμένο κύκλωμα (IC) και εκπέμπουν παλμούς ραδιοκυμάτων σε προκαθορισμένα χρονικά διαστήματα, τους οποίους συλλέγουν Access Points και στη βάση δεδομένων γίνεται η ανάλυση και εκτίμηση της θέσης τους. Οπότε, χρειάζονται την ίδια υποδομή με το Wi-Fi και τα BLEs συν τους ενεργούς πομπούς(beacons).

Τα passive RFIDs βασίζουν τη λειτουργία τους μέσω της επαγωγικής σύζευξης (Inductive coupling), παίρνουν την απαιτούμενη ενέργεια που χρειάζονται για τη λειτουργία τους από τον παλμό της κεραίας «αναγνώστη» και μετά εκπέμπουν τον κωδικό τους πίσω στην ίδια κεραία «αναγνώστη». Στα πλεονεκτήματά τους συγκαταλέγονται το μέγεθός τους και η φτηνή κατασκευή των tags. Βασικό μειονέκτημά τους είναι η μικρή απόσταση ανάγνωσης (γύρω στα 1 με 2 μέτρα), η οποία μαζί με την απλή αρχή λειτουργία τους καθιστά αναγκαία τη χρήση τους σε μεγάλους αριθμούς. Σε έρευνά τους, οι Zhao et al. χρησιμοποίησαν passive tags σε προϊόντα και μέσω αλγόριθμου iεραρχικής ομαδοποίησης κατάφεραν να προσδιορίσουν ποια προϊόντα βρίσκονταν σε ποια καλάθια μέσω της όμοιας κίνησής τους (trajectory analysis). Άλλα παρατήρησαν ότι η ακρίβεια έπεφτε με την παρουσία περισσότερων χρηστών (Zhao et al., 2018).



Bluetooth Low Energy Beacons

Τα beacons λειτουργούν με βάση το πρωτόκολλο Bluetooth Low Energy (BLE) , το οποίο είναι μια εξελιγμένη εκδοχή του κοινού Bluetooth. Είναι σχεδιασμένα για χαμηλή κατανάλωση σε επίπεδο ενέργειας και για μεταφορά λιγότερων δεδομένων, συνήθως 1-20% της μέσης κατανάλωσης ενέργειας από αυτή του Bluetooth. Όντας μικροσυσκευές, τα beacons τροφοδοτούνται κατεξοχήν από μπαταρίες μικρού μεγέθους (coin shell batteries). Αυτές οι συσκευές ανάλογα τη χρήση τους και την παραμετροποίησή τους έχουν μέσο όρο ζωής από μερικές ημέρες μέχρι και δεκατέσσερα χρόνια (Gomez, Oller, & Paradells, 2012). Σε άλλες έρευνες έχει παρατηρηθεί ότι η διάρκεια ζωής τους κυμαίνεται ένα με δυο χρόνια³. Οι συσκευές εγκαθίστανται κατά κανόνα χωροταξικά δηλαδή σε τοίχους ή στις οροφές των κτιρίων για καλύτερη ακρίβεια. Επιπρόσθετα, η ανιχνευσιμότητα τους φτάνει τα 75 μέτρα (Nilsson), χωρίς εμπόδια και ανάλογα με το επίπεδο ισχύος. Όσο μεγαλύτερη είναι η ισχύς, τόσο μεγαλύτερη είναι ακρίβεια που μπορεί να εντοπιστεί η θέση σε εσωτερικούς χώρους (ανάλογα και την τεχνική που χρησιμοποιείται π.χ. RSSI). Όπως και οι άλλες τεχνικές που χρησιμοποιούν ραδιοκύματα υπάρχει διαφοροποίηση ανάλογα με τις συνθήκες, τη θέση τους στο κατάστημα και τη μορφολογία ενός καταστήματος (Stavrou, Bardaki, Papakyriakopoulos, & Pramatari, 2019). Σε περιπτώσεις που μεταξύ beacon και Smart device μεσολαβεί τοίχος μπορεί να μειωθεί σημαντικά το εύρος εκπομπής σήματος και μεταλλικές κατασκευές όπως ράφια και κολώνες μπορούν είτε να απορροφήσουν ή να σκεδάσουν τα ραδιοκύματα και να δώσουν λανθασμένη μέτρηση.

NFC

Η NFC λειτουργεί μόνο μεταξύ 1-5 cm και καταναλώνει σχετικά μεγαλύτερη ισχύ. Οι παθητικές ετικέτες NFC μπορούν να αποφορτίζονται και να ενεργοποιούνται πάλι μόνο όταν υπάρχει ένα πεδίο NFC. Αυτό εξαλείφει το NFC από πολλές περιπτώσεις χρήσης .Το NFC είναι ιδανικό για τις ανέπαφες πληρωμές ως αντικαταστάτης των Qrcode και είναι πιθανό να ενσωματωθεί παράλληλα με τις άλλες τεχνολογίες που αναφέρονται παραπάνω και να χρησιμοποιηθεί σε LBS . Άλλα έχει πολύ περιορισμένες δυνατότητες στη χρήση για RTLS.

³ υπάρχουν και beacons που χρησιμοποιούν και συμβατικές πήγες ρεύματος π.χ. usb.

Άλλες τεχνολογίες που είναι άξιες αναφοράς είναι:

VLC

Η Visual Light Communication (VLC), μπορεί να χρησιμοποιηθεί ως αυτόνομη λύση ή ως προσθήκη σε άλλες (Wi-Fi, BLEs). Χρησιμοποιεί λάμπες φωτισμού Led και μεταδίδει πληροφορία με το να αναβοσβήνει το led σε υψηλή συχνότητα ξεχωριστή για κάθε συσκευή. Η ανάγνωση του σήματος γίνεται μέσω της κάμερας σ' ένα κινητό τηλέφωνο (Q. Wang et al., 2018).

Image Recognition

Οι σημερινές κινητές συσκευές είναι εξοπλισμένες με κάμερες. Ένα σύστημα εύρεσης θέσης βασισμένο στην επαυξημένη πραγματικότητα (augmented reality) συγκρίνει εικόνες που λαμβάνονται από την κάμερα με ένα δείγμα εικόνων αποθηκευμένων σε μια βάση δεδομένων (Al Delail et al., 2013).

Περαιτέρω ανάλυση των τεχνολογιών εύρεσης θέσης μέσα σε εσωτερικούς χώρους και σύγκρισή τους βρίσκεται στην τέλος της εργασία στο **Appendix B**

3

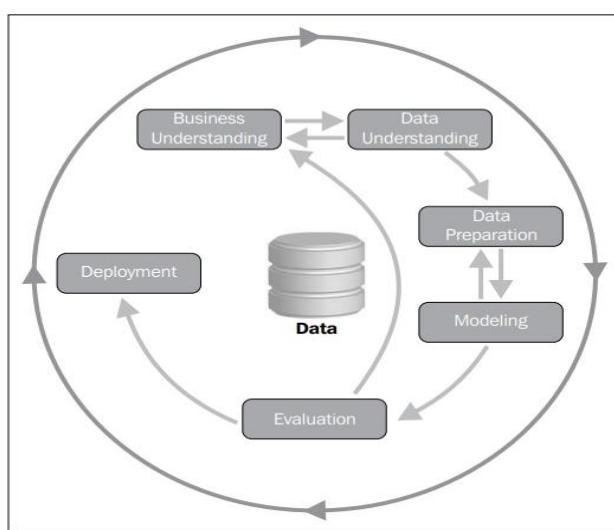
Μεθοδολογία

Σ' αυτό το κεφάλαιο παρουσιάζεται η μεθοδολογία που θα χρησιμοποιηθεί για την ανάλυση των δεδομένων καθώς και πως εφαρμόζεται στη συγκεκριμένη διπλωματική.

Στο δεύτερο υποκεφάλαιο περιγράφονται αναλυτικά όλες οι έννοιες που σχετίζονται με τον τρόπο κατάτμησης των καταναλωτών. Στη συνέχεια, παρατίθεται ο ορισμός της ανάλυσης συστάδων και περιγράφονται οι τεχνικές ομαδοποίησης που μπορούμε να έχουμε. Έπειτα, ορίζονται τα μέτρα εγγύτητας, δηλαδή τα μέτρα ομοιότητας και ανομοιότητας που χρησιμοποιούν οι διάφοροι αλγόριθμοι ομαδοποίησης εσωτερικά για την κατάτμηση των δεδομένων.

3.1 CRISP-DM Framework

Η εξαγωγή γνώσης από δεδομένα, ως μέρος της επιχειρηματικής ευφυΐας που χρησιμοποιεί μια επιχειρηση (ή ένας οργανισμός) για να βελτιώσει τις λειτουργίες της πρέπει να γίνεται με ένα μεθοδικό και σαφή τρόπο, με ορισμένα τεχνικά αλλά και οργανωτικά βήματα και όλα αυτά να εντάσσονται σε μια συνολική διεργασία. Γι' αυτόν το σκοπό δεν είναι αρκετό ο αναλυτής να έχει μόνο αντίληψη των επιχειρηματικών δεδομένων και καλή κατανόηση των μεθόδων, αλλά η όλη διαδικασία πρέπει να εντάσσεται σ' ένα καλώς ορισμένο σχέδιο και να έχει χαρακτηριστικά έργου (project), δηλαδή να έχει σαφώς καθορισμένους στόχους, προκαθορισμένα βήματα και να ακολουθεί πρότυπα οργάνωσης. Διάφορα μοντέλα έχουν προταθεί από τον επιχειρηματικό και ερευνητικό κόσμο (Shafique & Qaiser, 2014). Το πιο διαδεδομένο και συχνότερα χρησιμοποιούμενο μοντέλο (KDnuggets, 2007a) είναι το CRoss Industry Standard Process for DM (CRISP-DM) και περιλαμβάνει τον πιο ολοκληρωμένο οδηγό για τη διαδικασία εξόρυξης γνώσης από δεδομένα (KDnuggets, 2007b). Το CRISP-DM αναπτύχθηκε στα τέλη της δεκαετίας του 1990 (Chapman et al., 2000). Το κίνητρο για το έργο προέκυψε από το γεγονός ότι η εξόρυξη δεδομένων ήταν νέα στο οργανωτικό πλαίσιο, αναγκάζοντας τους αναλυτές να εφεύρουν τη διαδικασία ανάλογα με το έργο. Ο σχεδιασμός του CRISP-DM βασίστηκε σε μια σειρά εργαστηρίων (workshops) στην Ευρώπη και τις ΗΠΑ με μέλη πάνω από 200 επαγγελματίες του χώρου που ήταν μέλη αυτής της κοινοπραξίας (το cross-industry κατά λέξη σημαίνει ότι υπήρχε συμμετοχή διάφορων βιομηχανικών εταίρων). Στην παρακάτω εικόνα απεικονίζεται ένα υψηλού επιπέδου διάγραμμα του μοντέλου με τις 6 φάσεις που το αποτελούν, οι οποίες είναι: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation και Deployment .



Εικόνα 2: Παρουσιάζονται τα διαφορά στάδιά της μεθοδολογία CRISP-DM (Shearer, 2000)

Κάτι σημαντικό γι' αυτό το μοντέλο που μπορεί να παρατηρηθεί από την εικόνα είναι ότι το CRISP-DM θεωρεί ότι η εξόρυξη γνώσης δεν είναι μια διαδικασία γραμμικού τύπου ή μοντέλου καταρράκτη (waterfall) αλλά επαναληπτική.

Όπου:

- **Business Understanding** (Κατανόηση Επιχειρησιακού στόχου)

Είναι το πρώτο και πιο σημαντικό στάδιο γιατί αφορά το επιχειρησιακό πρόβλημα, τι επιδιώκει η επιχείρηση να επιλύσει, γι' αυτό το λόγο ο αναλυτής πρέπει να κατανοήσει από μια επιχειρηματική οπτική τι πραγματικά θέλει να πετύχει η επιχείρηση. Πιο συγκεκριμένα στο στάδιο αυτό γίνεται:

1) Καθορισμός των επιχειρησιακών στόχων :

Αρχικά, προσδιορίζονται οι ακριβείς επιχειρησιακοί στόχοι προς επίτευξη, ώστε η επιχείρηση να κατανοήσει με ακρίβεια τα προσδοκώμενα αποτελέσματα του project καθώς και τους πιθανούς περιορισμούς που μπορεί να προκύψουν και να θέσει τα ορθά κριτήρια της επιχειρησιακής επιτυχίας.

2) Περιγραφή της τρέχουσας κατάστασης, δηλαδή αναλύονται :

-οι περιορισμοί που έχει ο οργανισμός ως προς τους διαθέσιμους πόρους σε σχέση με τους αντίστοιχους που απαιτούνται για το συγκεκριμένο project

-οι διάφοροι παράγοντες που μπορούν να επηρεάσουν το project, τα ρίσκα και οι ευκαιρίες που μπορεί να υπάρξουν κ.α.

-τα κόστη και τα οφέλη της επιχείρησης σε σχέση με τη χρήση του συγκεκριμένου project

3) Περιγραφή των επιθυμητών αποτελεσμάτων του data mining project, ώστε να επιλυθεί το συγκεκριμένο επιχειρησιακό πρόβλημα υπό εξέταση και συγκεκριμένα, ο καθορισμός μιας επιθυμητής τιμής ελέγχου για την ακρίβεια του μοντέλου μας.

4) Δημιουργία του πλάνου που θα ακολουθηθεί για την εκτέλεση του project καθώς και αρχική αξιολόγηση των τεχνικών που θα χρησιμοποιηθούν για την εκτέλεσή του.

Το επιχειρησιακό ζήτημα που πραγματεύεται η παρούσα διπλωματική πηγάζει από την ανάγκη των λιανεμπορίων αρχικά να κατανοήσουν καλύτερα τη συμπεριφορά των πελατών στο φυσικό τους κατάστημα και μετέπειτα να την συγκρίνουν με τη συμπεριφορά των πελατών του ηλεκτρονικού τους καταστήματος. Μέσω καταλλήλων μετρικών να παρουσιαστούν σε επίπεδο συγκεντρωτικών αναφορών στη διοίκηση ώστε να την βοηθήσει στην λήψη αποφάσεων.



• Data Understanding (Κατανόηση Δεδομένων)

Το στάδιο αυτό περιλαμβάνει:

- 1) Εντοπισμός πηγών για άντληση δεδομένων, τρόπου άντλησης αυτών και αρχική εκτίμηση πιθανών προβλημάτων που μπορεί να εμφανιστούν
- 2) Την περιγραφή των δεδομένων που συλλέχθηκαν από τις παραπάνω πηγές ως προς την μορφοποίηση αυτών, το μέγεθος κ.α.
- 3) Την ανάλυση των δεδομένων (ανάλυση των μεταβλητών με χρήση βασικών στατιστικών εργαλείων) με προετοιμασία καθαρισμό και οπτικοποίηση δεδομένων και δημιουργία των σχετικών αναφορών
- 4) Τον έλεγχο της καταλληλότητας των δεδομένων (ως προς το format, πως θα διαχειριστούμε τα missing values, αν τα δεδομένα είναι επαρκή κ.α.)

Δηλαδή δεν έχει νόημα να τροφοδοτηθεί ένας μηχανισμός εξαγωγής γνώσης με ό,τι δεδομένα είναι διαθέσιμα χωρίς να έχει καθοριστεί ο στόχος προς διερεύνηση. Επίσης ο στόχος πρέπει να είναι εφικτός βάσει των δεδομένων που διατίθενται. Για την παρούσα εργασία επιλέχτηκε η μέθοδος της συσταδοποίησης των επισκέψεων για την επίλυση του επιχειρησιακού ζητήματος. Για την εν λόγω ανάλυση χρειάζονται δεδομένα που να δείχγουν τις περιοχές επίσκεψης του πελάτη προκειμένου να εξαχθεί το μονοπάτι που ακολούθησε.

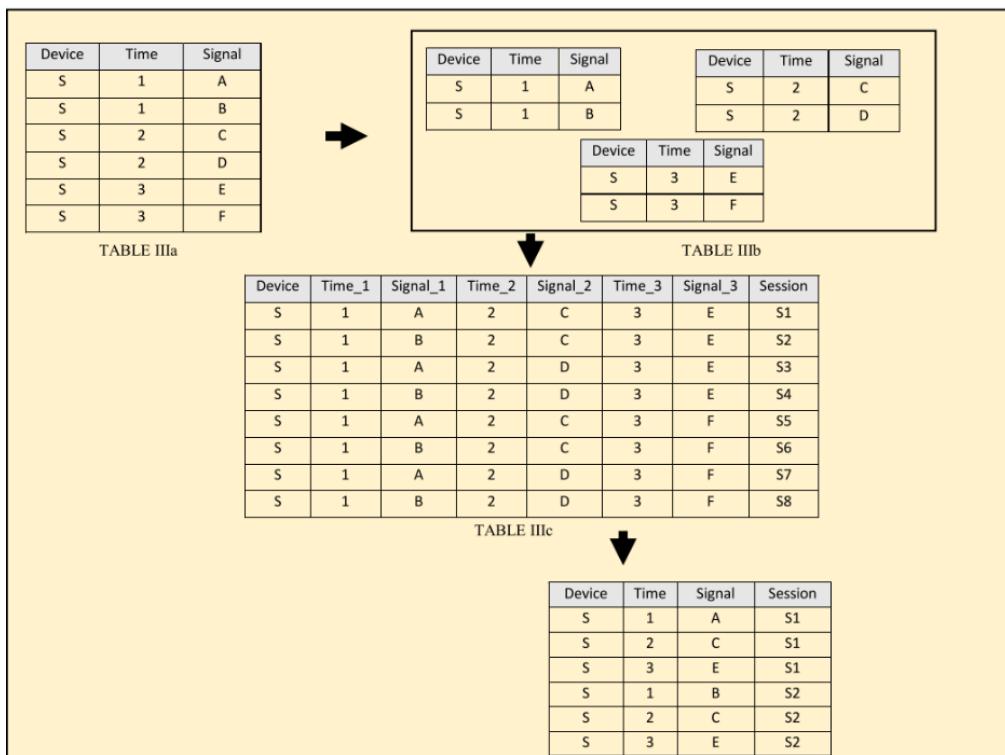
• Data Preparation (Προετοιμασία Δεδομένων)

Στο στάδιο αυτό γίνεται :

- 1) Επιλογή των αντληθέντων δεδομένων που τελικά θα χρησιμοποιηθούν
- 2) Διαχείριση των Missing values
- 3) Δημιουργία νέων μετρικών από τα υπάρχοντα δεδομένα για την καλύτερη ανάλυση του προβλήματος (μετασχηματισμός dummification μετατροπή δεδομένων από ordinal σε binary)
- 4) Μορφοποίηση και μετασχηματισμός των δεδομένων για τη διασφάλιση του data quality
- 5) Ενοποίηση των πινάκων δεδομένων, εκτέλεση aggregations, κ.α.

Ωστόσο, οι φάσεις προετοιμασίας και μοντελοποίησης των δεδομένων ροής (stream data) διαφέρουν από την παραδοσιακή εξόρυξη στατικών δεδομένων λόγω της χρονικής φύσης τους. Οι ερευνητές στο παρελθόν έχουν επικεντρωθεί σε μεγάλο βαθμό στην πρόταση των μοντέλων για τα δεδομένα. Ωστόσο, απαιτούνται νέες προσεγγίσεις για την προετοιμασία των δεδομένων

ροής ώστε αυτά να οργανωθούν κατάλληλα για τη μοντελοποίησή τους. Καθώς τα δεδομένα σε ακατέργαστη μορφή είναι δύσκολο να αναλυθούν, είναι ζωτικής σημασίας να μετατραπούν σε μια μορφή ενός δομημένου συνόλου δεδομένων ώστε να είναι σε θέση να εκτελεστεί οποιοδήποτε είδος ανάλυσης. Λόγω της προσθήκης της χρονικής διάστασης στην ήδη πολυνδιάστατη φύση των δεδομένων ροής, οι παραδοσιακές τεχνικές εξόρυξης δεδομένων ενδέχεται να μην είναι σε θέση να τα αναλύσουν ή να βγουν σωστά συμπεράσματα. (Kalgotra & Sharda, 2016)



Εικόνα 3: Μέθοδος επεξεργασία ακατέργαστων δεδομένων από την βάση δεδομένων (Kalgotra & Sharda, 2016)

Οι Kalgotra et al. διαπιστώνουν ότι τα δεδομένα από τους δέκτες ενός IPS αποθηκεύονται στη βάση με μη δομημένο τρόπο (unstructured) και πρέπει με καταλλήλους μετασχηματισμούς να ομαδοποιηθούν σε επίπεδο επίσκεψης (sessionizing).

Στη συνέχεια, για την παρούσα ανάλυση, τα «καθαρισμένα» δεδομένα για να τα πάρει ως είσοδο ο αλγόριθμος πρέπει να μετασχηματιστούν σε κατάλληλη μορφή, δηλαδή να δημιουργηθεί ένας πίνακας με τα δεδομένα, όπου κάθε γραμμή (εγγραφή) είναι μια επίσκεψη και κάθε μια στήλη (μετρική) είναι μια περιοχή του καταστήματος.

• Modeling (Μοντελοποίηση)

Στο στάδιο αυτό γίνεται:

- 1) Επιλογή του μοντέλου/ων που θα χρησιμοποιηθεί για την επύλυση του επιχειρησιακού μας προβλήματος (επιλογή του μοντέλου-αλγόριθμου και των σχετικών υποθέσεων υπό τις οποίες αυτός θα λειτουργήσει)
- 2) Σχεδιάζεται ο τρόπος εκτίμησης της απόδοσης του μοντέλου. Γίνεται χωρισμός των δεδομένων σε 2 datasets: το πρώτο set χρησιμοποιείται για την εκπαίδευση του μοντέλου μας στα δεδομένα του προβλήματος (training dataset) κα το δεύτερο set χρησιμοποιείται για την εκτίμηση της απόδοσης του μοντέλου
- 3) Γίνεται η παραμετροποίηση του μοντέλου βάσει των υποθέσεων λειτουργίας του μοντέλου (ρύθμιση των παραμέτρων του μοντέλου), για να βρεθεί η καλύτερη δυνατή τιμή ακρίβειας του μοντέλου
- 4) Γίνεται εκτίμηση του μοντέλου και πραγματοποιούνται οι κατάλληλες αλλαγές, για να πάρουμε ικανοποιητικά αποτελέσματα από το μοντέλο που δημιουργήσαμε (π.χ. αλλαγές μέχρι να φτάσουμε σε μια ικανοποιητική τιμή για το accuracy του μοντέλου)

Αφού επιλεγούν ο/οι αλγόριθμο/οι, στην προκειμένη περίπτωση αλγόριθμοι ομαδοποίησης (K-mode, K-medoids) πρέπει να ελεγχτούν με κατάλληλα εργαλεία για να παραμετρικοποιηθεί ο κάθε αλγόριθμος ώστε να βρεθεί το βέλτιστο αποτέλεσμα του (για την παραπάνω διαδικασία γίνεται εκτενής αναφορά στην επομένη παράγραφο).

• Evaluation (Αξιολόγηση)

Μετά την κατασκευή ενός μοντέλου που φαίνεται ακριβές βάσει των τεχνικών αξιολόγησης της μοντελοποίησης που σχεδιάστηκαν στο προηγούμενο βήμα, τα αποτελέσματα θα πρέπει να αξιολογηθούν και από τη σκοπιά των επιχειρησιακών στόχων. Σκοπός είναι να καθοριστεί η χρησιμότητα της εξαχθείσας γνώσης και η ικανότητά της να λύνει το πρόβλημα που είχε τεθεί αρχικά. Ένα μοντέλο πρόβλεψης μπορεί να αποκαλύπτει είτε προφανείς κανόνες, είτε κανόνες που είναι περίπλοκοι ή δεν έχουν κάποια ρεαλιστική ερμηνεία και άρα καμία αξία για την επιχείρηση. Στην περίπτωση που το μοντέλο δεν έχει αξία για την επιχείρηση, αυτό μπορεί να οφείλεται είτε στα δεδομένα που έχουν χρησιμοποιηθεί (οπότε θα πρέπει να επανεπιλεγούν και μετασχηματιστούν), ή το πρόβλημα δεν είναι σαφώς ορισμένο και θα πρέπει να γίνει επιστροφή στο στάδιο του Business Understanding και να επαναπροσδιοριστεί το πρόβλημα.

Σ' αυτό το βήμα, εν προκειμένω στην παρούσα εργασία, αξιολογούνται τα αποτελέσματα που έβγαλε ο αλγόριθμος και γίνεται η ερμηνεία τους για πιθανή επιχειρησιακή χρήση. Αξιολογείται κατά πόσο επιτεύχθηκε ο στόχος που είχε τεθεί να βρεθούν οι αγοραστικές αποστολές των πελατών (για ποιο λόγο επισκέπτονται το κατάστημα).



- **Deployment** (Ανάπτυξη)

Στο στάδιο αυτό:

1. Η γνώση που αποκομίσθηκε από την όλη διαδικασία, θα πρέπει να οργανωθεί σε κατάλληλη μορφή και να γίνει συγγραφή της τελικής αναφοράς, η οποία θα αξιοποιηθεί από τη διοίκηση της επιχείρησης.
2. Δημιουργείται ένα πλάνο παρακολούθησης και συντήρησης του μοντέλου (training σε νέα δεδομένα κλπ.). Το μοντέλο θα πρέπει να είναι διαθέσιμο σε λειτουργική μορφή, δηλαδή να μπορεί να παίρνει νέα δεδομένα εισόδου και να τα ερμηνεύει βάσει των κανόνων που έχει «μάθει» από την προηγούμενη εκπαίδευση.
3. Παράγεται η τελική αναφορά στην οποία περιγράφεται η όλη διαδικασία για την ανάπτυξη του μοντέλου, καθώς και οι αδυναμίες και περιορισμοί λειτουργίας του μοντέλου.

Business Understanding	Data Understanding	Data Preparation	Modeling	Evaluation	Deployment
<p>Determine Business Objectives Background Business Objectives Business Success Criteria</p> <p>Assess Situation Inventory of Resources Requirements, Assumptions, and Constraints Risks and Contingencies Terminology Costs and Benefits</p> <p>Determine Data Mining Goals Data Mining Goals Data Mining Success Criteria</p> <p>Produce Project Plan Project Plan Initial Assessment of Tools and Techniques</p>	<p>Collect Initial Data Initial Data Collection Report</p> <p>Describe Data Data Description Report</p> <p>Explore Data Data Exploration Report</p> <p>Verify Data Quality Data Quality Report</p>	<p>Select Data Rationale for Inclusion/Exclusion</p> <p>Clean Data Data Cleaning Report</p> <p>Construct Data Derived Attributes Generated Records</p> <p>Integrate Data Merged Data</p> <p>Format Data Reformatted Data</p> <p>Dataset Dataset Description</p>	<p>Select Modeling Techniques Modeling Technique Assumptions</p> <p>Generate Test Design Test Design</p> <p>Build Model Parameter Settings Models Model Descriptions</p> <p>Assess Model Model Assessment Revised Parameter Settings</p>	<p>Evaluate Results Assessment of Data Mining Results w.r.t. Business Success Criteria Approved Models</p> <p>Review Process Review of Process</p> <p>Determine Next Steps List of Possible Actions Decision</p>	<p>Plan Deployment Deployment Plan</p> <p>Plan Monitoring and Maintenance Monitoring and Maintenance Plan</p> <p>Produce Final Report Final Report Final Presentation</p> <p>Review Project Experience Documentation</p>

Εικόνα 4: Συνοπτική παρουσίαση και εξήγηση των στόχων σε κάθε βήμα της μεθοδολογία (Chapman et al., 2000)

3.2 Ανάλυση Συστάδων (*Cluster Analysis*)

Η ανάλυση συστάδων είναι μια από τις συχνότερα χρησιμοποιούμενες μεθόδους στην ανάλυση δεδομένων προκειμένου να διερευνηθούν οι σχέσεις ανάμεσα στα δεδομένα και να εξαχθεί κάποιο αρχικό αποτέλεσμα. Σαν μέθοδος έχει στόχο να διαχωρίσει τις υπάρχουσες παρατηρήσεις (σύνολο των παρατηρήσεων) βάσει της πληροφορίας που υπάρχει στα εκάστοτε γνωρίσματα των δεδομένων σε επιμέρους υποσύνολα (συστάδες). Δηλαδή η ανάλυση συστάδων εξετάζει πόσο όμοιες είναι οι παρατηρήσεις ως προς τον αριθμό των μεταβλητών-χαρακτηριστικών (features) που έχουν οριστεί ως κριτήρια (Huang Z., 1998).

Βασικό γνώρισμα της μεθόδου είναι το γεγονός ότι δεν υπάρχει εκ των προτέρων γνώση σχετικά με την ύπαρξη ομάδων, για το λόγο αυτό η ανάλυση συστάδων ανήκει στην κατηγορία της μη επιβλεπόμενης μάθησης (unsupervised learning). Ουσιαστικά μπορεί να επεξεργαστεί δεδομένα στα οποία δεν υπάρχει κάποια στήλη-στόχος από την οποία, να μπορεί να «διδαχτεί» για την ομαδοποίηση των δεδομένων (Γ. KYPKOΣ, 2015).

Οι αλγόριθμοι της ανάλυσης συστάδων διακρίνονται σε ιεραρχικούς (hierarchical), σε διαμεριστικούς (partitional), σε αλγόριθμους που βασίζονται στην πυκνότητα και σε πιθανοθεωρητικά μοντέλα (model based clustering). Η ιεραρχική συσταδοποίηση δημιουργεί μια ιεραρχία από συστάδες που δημιουργούνται βαθμιαία. Οι διαιρετικοί, όπως δηλώνει και το όνομά τους, διαιρούν ένα σύνολο ή αντικειμένων σε συστάδες. Απ' την άλλη, η διαμεριστική συσταδοποίηση καταλήγει σ' ένα σύνολο ή συστάδων, όπου κάθε στοιχείο δεδομένων ανήκει σε μια συστάδα, στην οποία αντιστοιχεί ένα κέντρο (centroid), που αντιπροσωπεύει το σύνολο των στοιχείων της συγκεκριμένης συστάδας. Αλγόριθμοι αυτής της κατηγορίας είναι οι K-Means, K-Modes, K-Medoids (ή ορθότερα Partitioning Around Medoids). Οι βασιζόμενοι στην πυκνότητα αλγόριθμοι είναι ικανοί να δημιουργήσουν συστάδες τυχαίων μορφών (σε σχέση με τους προηγούμενους τύπους που σχηματίζουν γεωμετρικά σχήματα, για παράδειγμα ο K-Means σχηματίζει ελλειψοειδή). Αυτού του τύπου οι αλγόριθμοι (DBSCAN, OPTICS) ομαδοποιούν τα στοιχεία με βάση το πλήθος των δεδομένων που υπάρχουν γύρω από μια περιοχή και συνεχίζουν να αυξάνονται όσο ο αριθμός των αντικειμένων υπερβαίνει κάποια ορισμένη τιμή (αναθέτει ένα αντικείμενο σε μια συστάδα βάσει της πυκνότητας των αντικειμένων στη συγκεκριμένη περιοχή). Πλεονέκτημα της μεθόδου αυτής είναι ότι παράγει τυχαίες μορφές και ότι είναι ανθεκτική σε ακραίες τιμές των μεταβλητών. Τέλος, οι αλγόριθμοι βασισμένοι σε πιθανοθεωρητικά μοντέλα κατατάσσουν τα στοιχεία σε συστάδες βάσει την πιθανότητας να ανήκουν σε μια κατανομή. Μ' ένα τέτοιο μοντέλο ο αναλυτής έχει τη δυνατότητα να προβεί σε στατιστική συμπερασματολογία, εφόσον έχει κάνει πρώτα κάποιες υποθέσεις για τα δεδομένα, π.χ. ότι ακολουθούν κάποια κατανομή κτλ. Παράδειγμα αυτής της κατηγορίας είναι ο EM (Expectation Maximization) αλγόριθμος.



Η ομοιότητα (Maronna, 2016) μεταξύ δύο αντικειμένων είναι ένας αριθμός που μετράει το βαθμό στον οποίο δύο αντικείμενα είναι όμοια. Συνεπώς, οι ομοιότητες είναι υψηλότερες για ζεύγη αντικειμένων τα οποία μοιάζουν περισσότερο. Οι ομοιότητες είναι μη αρνητικοί αριθμοί και συχνά βρίσκονται ανάμεσα στο 0 (καμία ομοιότητα) και στο 1 (απόλυτη ομοιότητα). Αντίθετα η ανομοιότητα μεταξύ δύο αντικειμένων είναι ένα αριθμητικό μέτρο του βαθμού στον οποίο δύο αντικείμενα είναι ανόμοια. Συνεπώς, οι ομοιότητες είναι ψηλότερες για περισσότερο όμοια ζεύγη αντικειμένων. Οι ανομοιότητες εκφράζονται στο διάστημα [0,1], αλλά είναι πιθανό, λόγω του ορισμού τους να κυμαίνονται μεταξύ του 0 και του ∞ .

Ο όρος εγγύτητα (Maronna, 2016) αναφέρεται είτε στην ομοιότητα είτε στην ανομοιότητα. Η ομοιότητα μεταξύ δύο αντικειμένων (παρατηρήσεων) είναι μια συνάρτηση της εγγύτητας μεταξύ των αντίστοιχων χαρακτηριστικών τους. Στην περίπτωση που αναλύονται πυκνά δεδομένα, όπως είναι οι χρονοσειρές ή τα πολυδιάστατα σημεία (αντικείμενα με παραπάνω από 2 μετρικές), το μέτρο εγγύτητας που χρησιμοποιείται συνήθως είναι η Ευκλείδεια απόσταση (στη γενικευμένη της μορφή). Αντίθετα, όταν εξετάζονται σποραδικά δεδομένα, όπως είναι δεδομένα καλαθιού αγορών, τότε τα πιο συχνά χρησιμοποιούμενα μέτρα ομοιότητας είναι ο συντελεστής Jaccard και η ομοιότητα συνημίτονου (Karlis, 2005).

Ένας τρόπος καθορισμού του βαθμού ομοιότητας δύο παρατηρήσεων είναι με τη χρήση της απόστασής τους. Οι παρατηρήσεις ορίζονται ως σημεία σε έναν πολυδιάστατο χώρο. Η απόστασή τους σ' αυτόν τον χώρο αποτελεί το μέτρο της ομοιότητάς τους. Εάν όλα τα γνωρίσματα είναι αριθμητικά, τότε για τον υπολογισμό της ανομοιότητας χρησιμοποιείται η Ευκλείδεια απόσταση ή κάποια παραλλαγή της, όπως η απόσταση Manhattan ή η απόσταση Minkowski.

3.2.1 Μέτρα Απόστασης Data Preparation – Format Data -

Τα τρία βασικά μέτρα αποστάσεων γενικεύονται από το μέτρο απόστασης Minkowski που δίνεται από την παρακάτω εξίσωση(Karlis, 2005).

$$d(x, y) = \left(\sum_{i=1}^n (|x_i - y_i|)^r \right)^{1/r}$$

Όπου n είναι οι μεταβλητές που έχουμε και r είναι μια παράμετρος που καθορίζει τις διαστάσεις του μέτρου απόστασης.

Παρακάτω δίνονται τα πιο κοινά παραδείγματα των αποστάσεων Minkowski.

Για r ισούται με 1: είναι η Manhattan απόσταση στην βιβλιογραφία αναφέρεται και City Block.



Για τις ισούται με 2: είναι η Ευκλείδεια απόσταση.

Άλλο μέτρο ομοιότητας μεταξύ αντικειμένων που περιέχουν μόνο δυαδικά χαρακτηριστικά είναι ο συντελεστής **Jaccard** (Jaccard Coefficient): Χρησιμοποιείται συχνά λόγω της ιδιότητάς του να μετρά την ομοιότητα δύο αντικειμένων που αποτελούνται από ζευγάρια καταστάσεων που δεν είναι ισότιμες μεταξύ τους. Οι καταστάσεις παίρνουν μόνο δύο τιμές 0 και 1 και συμβολίζουν διαφορετικής αξίας πληροφορία. Με άλλα λόγια, αυτός ο συντελεστής επιλέγεται γιατί απεικονίζει ασύμμετρες μεταβλητές που μια εκ των δύο τιμών δεν είναι χρήσιμη για την εξαγωγή συμπερασμάτων.

3.2.2 Αλγόριθμοι Modeling

K-Means

Η μέθοδος αυτή προσπαθεί να εντοπίσει παρόμοιες ομάδες αντικείμενων βάσει επιλεγμένων χαρακτηριστικών (MacQueen, 1967). Όπως και στις περισσότερες τεχνικές κατηγοριοποίησης έτσι και στην K-Means ανάλυση, ο αναλυτής καθορίζει τον επιθυμητό αριθμό ομάδων ή τμημάτων. Κατά τη διάρκεια της διαδικασίας υπολογίζεται η απόσταση του κάθε αντικειμένου, στην περίπτωση μας, ο πελάτης από τα κέντρα βάρους των συστάδων (centroids). Η διαδικασία επαναλαμβάνεται μέχρι να μεγιστοποιηθεί η απόσταση μεταξύ των κέντρων των συστάδων ή να επιτευχθεί κάποιο άλλο καθορισμένο κριτήριο. Οι πελάτες τοποθετούνται στη συστάδα με το πλησιέστερο γεωμετρικό κέντρο. Ο αλγόριθμος χρησιμοποιεί τα δεδομένα που μπορούν να παρέχουν πληροφορίες σχετικά με τη δυνατότητα κάθε μεταβλητής να διαφοροποιήσει αυτά τα τμήματα. Ένα από τα προτερήματα του αλγόριθμου K-Means είναι η ευκολία του να εκτελεστεί καθώς και η ταχύτητα εκτέλεσής του. Ως δεδομένα εισόδου χρησιμοποιούνται τα αποτελέσματα της ανάλυσης παραγόντων ή/και της ανάλυσης RFM (μεταβλητές για την αξιολόγηση του πελάτη). Η ανάλυση συστάδων με τον K-Means περιγράφει την πολυδιάστατη συμπεριφορά των πελατών αναδεικνύοντας ομάδες πελατών που τους αρέσουν ή όχι (ανάλογα με τις τιμές των χαρακτηριστικών) κοινές κατηγορίες.

Ο βασικός αλγόριθμος K-Means λειτουργεί ως εξής:

Βήμα 1. Είσοδος: λαμβάνει ως παράμετρο εισόδου από τον αναλυτή, το πλήθος των επιθυμητών συστάδων k και τυχαία επιλέγει αντίστοιχο αριθμό αντικειμένων, κάθε ένα από τα οποία αναπαριστά το κέντρο κάθε συστάδας για την πρώτη επανάληψη.

Βήμα 2. Κάθε ένα αντικείμενο καταχωρείται στην συστάδα εκείνη με την οποία είναι πιο όμοιο, με βάση την απόσταση μεταξύ του αντικειμένου και του κέντρου της συστάδας.

Βήμα 3. Πριν ολοκληρωθεί η επανάληψη τα κέντρα των συστάδων επανυπολογίζονται με βάση τα σημεία που έχουν συσχετιστεί με τη συγκεκριμένη συστάδα.

Βήμα 4. Η διαδικασία αυτή επαναλαμβάνεται μέχρι να μην υπάρχει σημείο που να αλλάζει συστάδα ή μέχρι τα κέντρα βάρους (centroids) να μένουν σταθερά.

Ωστόσο η τεχνική αυτή δεν μπορεί να διαχειριστεί μεταβλητές που έχουν διαφορετικές κλίμακες και χρειάζεται να γίνει κανονικοποίηση των τιμών. Επίσης, λόγω του ότι υπολογίζει τα κέντρα βάρους επηρεάζεται από τις ακραίες τιμές των δεδομένων (outliers).

K-Modes

Ο αλγόριθμος K-Modes είναι παραλλαγή του K-Means καθώς είναι κατάλληλος να επεξεργάζεται δυαδικά δεδομένα και έχει προταθεί από τον Huang (Huang, 1998; Huang, 1997). Η μέθοδος αυτή προσπαθεί να εντοπίσει παρόμοιες ομάδες αντικείμενων βάσει επιλεγμένων χαρακτηριστικών τα οποία είναι σε δυαδική μορφή και χρειάζεται να της οριστεί ο αριθμός των συστάδων. Ο αλγόριθμος για να μετρήσει την ομοιότητα, χρησιμοποιεί την απόσταση των αντικείμενων από τις συστάδες με μέτρα απόστασης όπως η Jaccard και η simple matching, ακολουθώντας τα παρακάτω βήματα :

Βήμα 1. Αφού επιλέξει τα κέντρα των συστάδων αναθέτει σε αυτές τα πιο όμοια αντικείμενα.

Βήμα 2. Πριν ολοκληρωθεί η επανάληψη, τα κέντρα των συστάδων επανυπολογίζονται βάσει της επικρατούσας τιμής (mode) των αντικείμενων που έχουν συσχετιστεί με τη συγκεκριμένη συστάδα.

Βήμα 3.Η διαδικασία αυτή επαναλαμβάνεται μέχρι να μην υπάρχει σημείο που να αλλάζει συστάδα ή μέχρι κάποιο κριτήριο τερματισμού (πχ N αριθμός επαναλήψεων).

Ιεραρχική ομαδοποίηση

Η ιεραρχική ομαδοποίηση είναι μια άλλη μέθοδος της ανάλυσης συστάδων. Οι μέθοδοι λειτουργούν ιεραρχικά, κάθε επανάληψη του αλγορίθμου συνίσταται σε μια διαδικασία διαδοχικών συγχωνεύσεων ή διασπάσεων των συστάδων. Οι σχετικές τεχνικές αντιστοίχως χωρίζονται σε συσσωρευτικές (agglomerative) και διαιρετικές (divisive).

Οι συσσωρευτικές (agglomerative) μέθοδοι αρχικά θεωρούν κάθε ξεχωριστή παρατήρηση ως μια συστάδα. Οι πιο όμοιες παρατηρήσεις επιλέγονται και συγχωνεύονται, δημιουργώντας μια



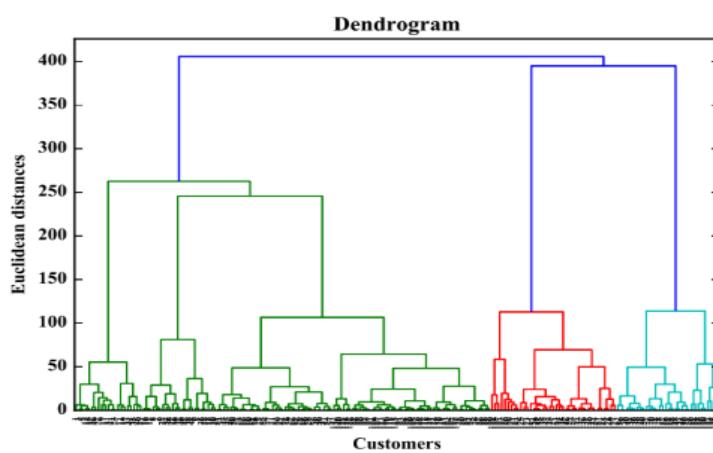
νέα συστάδα. Από τις συστάδες που προκύπτουν, επιλέγονται οι πιο όμοιες και συγχωνεύονται. Η διαδικασία επαναλαμβάνεται μέχρι να ενταχθούν όλα τα αντικείμενα σε μια ενιαία συστάδα. Η μέθοδος ξεκινάει από το κατώτατο επίπεδο της ιεραρχίας (όλες οι παρατηρήσεις θεωρούνται συστάδες) και με διαδοχικές συγχωνεύσεις σταδιακά ανέρχονται τα επίπεδα. Υιοθετούν δηλαδή μια προσέγγιση «από κάτω προς τα επάνω» (bottom up approach) και θεωρούνται οι πιο διαδεδομένες (Karlis, 2005). Αντίστροφα λειτουργούν οι διαιρετικές.

Βήμα 1. Αρχικά, δημιουργείται ο πίνακας αποστάσεων και καταγράφονται οι αποστάσεις μεταξύ των σημείων. Κάθε ένα από τα N αντικείμενα θεωρείται ως μια ξεχωριστή συστάδα.

Βήμα 2. Βάσει του πίνακα αποστάσεων βρίσκεται η μικρότερη τιμή, η οποία είναι η μικρότερη απόσταση $d(X_i, X_j)$ ανάμεσα σε δύο αντικείμενα X_i, X_j . Τότε συνενώνονται τα δύο αντικείμενα σε μια συστάδα.

Βήμα 3. Επαναλαμβάνεται το βήμα 2, $N-1$ φορές βάσει της μεθόδου υπολογισμού της απόστασης. Σε κάθε επανάληψη καταγράφονται οι συστάδες που συγχωνεύονται. Η μικρότερη απόσταση μπορεί να αφορά ένα αντικείμενο και μια συστάδα ή 2 συστάδες.

Δυο βασικές διαφορές της Ιεραρχικής Ομαδοποίησης με τις Διαχωριστικές Μεθόδους (K-Modes, K-Means) είναι: Πρώτον, κάθε ενέργεια (ενώσεις για τις συσσωρευτικές και διασπάσεις για τις διαιρετικές), η οποία πραγματοποιείται σε μια επανάληψη του αλγόριθμου, δεν είναι αντιστρέψιμη. Από τη στιγμή που δύο αντικείμενα ενταχθούν στην ίδια ομάδα, παραμένουν στην ίδια ομάδα χωρίς να μπορεί να επανεξεταστεί η ομοιότητά τους. Δεύτερον, δεν χρειάζεται να οριστεί ο αριθμός των συστάδων εκ των προτέρων, κάτι το οποίο απαιτεί διερεύνηση στους διαχωριστικούς αλγορίθμους. Παρακάτω φαίνεται το δενδρογραμμά που δείχνει όλα τα βήματα του αλγορίθμου (πως έγιναν οι ενοποίησης των συστάδων) και φαίνεται ότι τα δεδομένα μπορούν να χωριστούν σε 5 χαρακτηριστικές συστάδες.



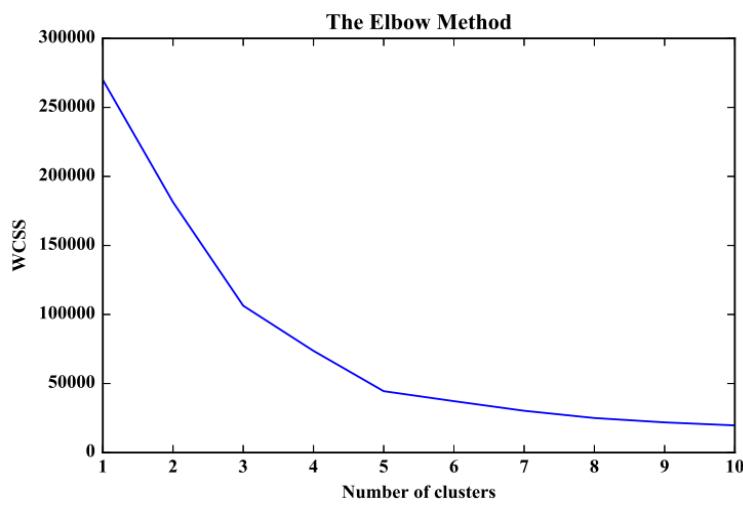
Εικόνα 5: Παράδειγμα δενδρογράμματος από (Tripathi et al., 2018)

3.2.3 Επιλογή αριθμού συστάδων πελατών Build Model

Βασικό μειονέκτημα των διαιρετικών αλγόριθμων που ήδη αναφέρθηκε, είναι ότι πρέπει να οριστεί ο αριθμός των συστάδων από το χρήστη/ερευνητή. Για αυτό τον λόγο, έχουν αναπτυχθεί διάφορες μέθοδοι.

Elbow method

Η μέθοδος «elbow» είναι μια από τις παλαιότερες και συνήθεις μεθόδους (Kodinariya & Makwana, 2013). Καταλήγει σε μια γραφική απεικόνιση των αριθμών των συστάδων στον ένα άξονα και του κόστους (cost function) στον άλλο (Syakur, Khotimah, Rochman, & Dwi Satoto, 2018). Στην περίπτωση του K-Means ως κόστος εννοείται η τετραγωνική απόσταση του κάθε αντικειμένου από το κεντροειδές, το κέντρο της κάθε συστάδας στην οποία ανήκει Sum of Squared Error⁴. Η βασική ιδέα είναι ότι όσο αυξάνεται ο αριθμός των συστάδων k, τόσο μειώνεται το κόστος (άθροισμα των αποστάσεων των σημείων από τα κέντρα) γιατί δημιουργούνται νέα κέντρα και μικραίνουν οι αποστάσεις των σημείων από αυτά. Ως «βέλτιστο» k ορίζεται το σημείο που αλλάζει ο ρυθμός μείωσης του SSE, δηλαδή τα σημεία καμπής της γραφικής. Για μεγάλο αριθμό k η γραφική παράσταση τείνει ασυμπτωτικά τον οριζόντιο άξονα. Ένα σημείο καμπής k αποτελεί τοπικό βέλτιστο της λύσης καθώς δείχνει ότι για τον επόμενο αριθμό k+1 συστάδων πέφτει ο ρυθμός της εξήγησης της διακύμανσης όπως από τον k-1 στον k. Τα ακραία σημεία καμπής κάνουν την γραφική παράσταση να μοιάζει με «αγκώνα» για αυτό ονομάστηκε και elbow.



Εικόνα 6: Παράδειγμα elbow οπού το βέλτιστο k=5 είναι το σημείο καμπής της γραφικής (Tripathi et al., 2018)

⁴βιβλιογραφικά αναφέρεται ως Within Cluster Sum of Squares (WCSS) (Aggarwal & Reddy, 2014; Tripathi et al., 2018)

$$SSE = \sum_{k=1}^K \sum_{x \in C_k} d(x, c_k)^2$$

Από πάνω δίνεται η εξίσωση υπολογισμού της SSE , όπου k ο αριθμός των συστάδων και x τα αντικείμενα που ανήκουν συγκεκριμένη συστάδα Ci.

Το μειονέκτημα αυτής της μεθόδου είναι ότι δεν είναι πάντα εύκολο να αναγνωριστεί το σημείο αλλαγής του ρυθμού μείωσης του κόστους, διότι σ' ένα γράφημα μπορεί να περιέχονται παραπάνω από ένα τέτοια σημεία ή να μην είναι ευκρινή.

Silhouette score

Η average Silhouette score είναι άλλη μια μέθοδος εύρεσης του βέλτιστου αριθμού συστάδων. Αυτό που κάνει είναι να μετρά την απόσταση κάθε παρατήρησης με τις παρατηρήσεις που ανήκουν στην πλησιέστερη σε αυτή συστάδα και μετά συγκρίνει την απόσταση που έχει με τις παρατηρήσεις στη συστάδα που ανήκει. (Nies et al., 2019; Tripathi et al., 2018). Μ' αυτό τον τρόπο βρίσκει κατά πόσο μια παρατήρηση είναι σωστά τοποθετημένη στη συστάδα που την έχει «αναθέσει» ο αλγόριθμος.

$$S_i = \frac{b_i - a_i}{\max\{a_i, b_i\}}$$

Και οι όροι b_i , a_i υπολογίζονται ως εξής:

$$a_i = \frac{1}{|C_i|} \sum_{j \in C_i} d(i, j) \quad b_i = \min_{c_k: k \neq i} \sum_{j \in C_i} \frac{d(i, j)}{|C_i|}$$

Όπου a_i συμβολίζει τη μέση απόσταση μεταξύ της παρατήρησης i και των παρατηρήσεων που ανήκουν στην ίδια συστάδα. Ενώ το b_i συμβολίζει τη μέση απόσταση μεταξύ της παρατήρησης i και των παρατηρήσεων της πλησιέστερης συστάδας.

Τέλος το average Silhouette score υπολογίζεται ως ο μέσος όρος των αποστάσεων

$$\frac{1}{N} \sum_{i=1}^N S_i$$

Οι τιμές του δείκτη Silhouette score κυμαίνονται μεταξύ $[-1, 1]$ όπου το 1 συμβολίζει την τέλεια αντιστοίχιση, δηλαδή το στοιχείο i ανήκει στη συστάδα.

4

Case Study

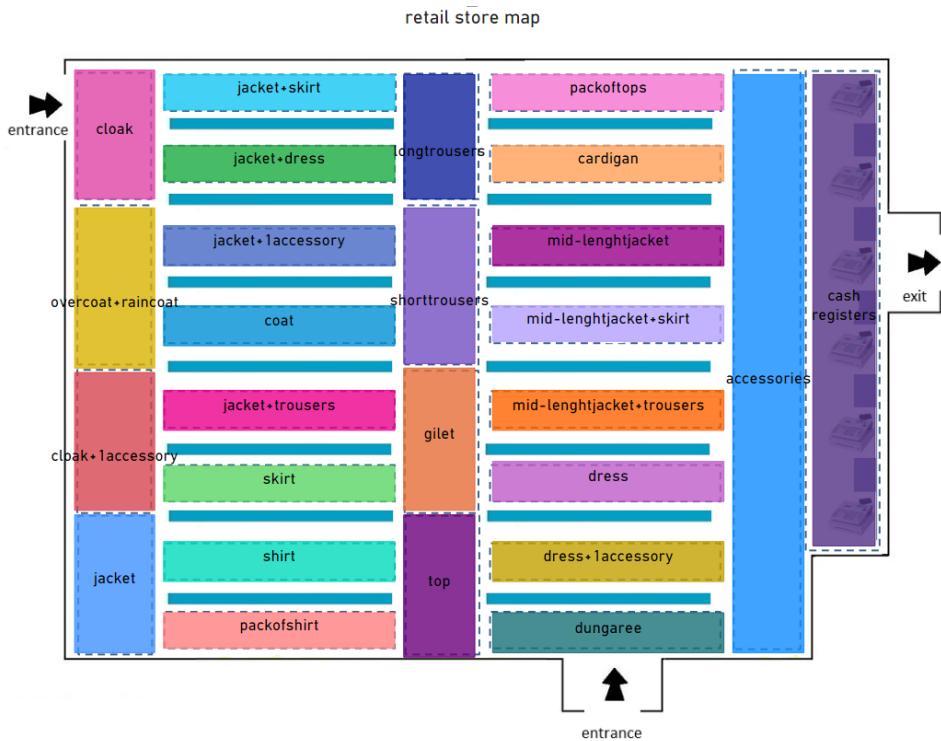
Στο παρόν κεφάλαιο, θα εξεταστεί ένα πρακτικό παράδειγμα ανάλυσης της καταναλωτικής συμπεριφοράς των πελατών, εντός ενός καταστήματος λιανικού εμπορίου ρουχισμού, με εφαρμογή της μεθόδου CRISP-DM.

Όπως έχει αναφερθεί και στην εισαγωγή, στόχος της εργασίας αυτής είναι να μελετηθεί η χρήση χωροχρονικών δεδομένων για την ανάλυση της συμπεριφοράς των πελάτων σ' ένα φυσικό κατάστημα και να βρεθούν κατάλληλες μετρικές για αυτό το σκοπό προκειμένου να στηριχθεί πιο αποτελεσματικά η διαδικασία λήψης αποφάσεων πάνω σε τέτοια δεδομένα. Η μεθοδολογία που ακολουθεί η εργασία για την ανάλυση των δεδομένων είναι η CRISP-DM στην οποία αναφερθήκαμε στο προηγούμενο κεφάλαιο.

Με βάση το συγκεκριμένο παράδειγμα, μας ζητήθηκε η δημιουργία των κατάλληλων εργαλείων για τη μέτρηση και μελέτη της επισκεψιμότητας των πελατών στις διάφορες περιοχές του καταστήματος, να ανακαλυφθούν οι περιοχές υψηλής επισκεψιμότητας, δηλαδή τα σημεία στα οποία προτιμούν να παραμένουν για περισσότερο χρόνο οι πελάτες εντός του συγκεκριμένου καταστήματος (hot spots), όπως και οι αντίστοιχες περιοχές χαμηλής επισκεψιμότητας (dead spots).

Επιπλέον μας ζητήθηκε ο χαρακτηρισμός της αγοραστικής συμπεριφορά των πελατών, δηλαδή η ανεύρεση των αγοραστικών αποστολών (shopping missions) αυτών (ποιος είναι δηλαδή ο συγκεκριμένος στόχος για τον οποίο επισκέπτονται οι πελάτες το κατάστημα, αν υπάρχει κάποιος συγκεκριμένος συνδυασμός προϊόντων ιματισμού στην απόκτηση των οποίων αυτοί αποσκοπούν κ.λπ.).

Στο υποθετικό παράδειγμά μας, η διαμόρφωση των διαφόρων περιοχών εντός αυτού, ανάλογα με τις διαφορετικές κατηγορίες ειδών ιματισμού, παρουσιάζει την παρακάτω εικόνα, στην οποία φαίνεται ότι το κατάστημά μας ακολουθεί το πρότυπο διαμόρφωσης τύπου grid (grid layout), δηλαδή διαθέτει κάθετους και παράλληλους διαδρόμους.



Εικόνα 7: Θέση των κατηγοριών στο κατάστημα

Με βάση τα παραπάνω: Πρώτον θα παρουσιαστούν τα δεδομένα που χρησιμοποιήθηκαν στην εργασία στο πλαίσιο της Κατανόησης των Δεδομένων. Δεύτερον για το βήμα Προετοιμασία Δεδομένων θα γίνουν μετατροπές στα δεδομένα για να έρθουν στην κατάλληλη μορφή για επεξεργασία. Θα δημιουργηθούν νέες στήλες, θα ενωθούν πίνακες για να ταυτιστούν τα χαρακτηριστικά των εγγραφών και θα γίνει διαχείριση των ελλιπών εγγραφών.

Θα γίνουν οπτικοποιήσεις των μετρικών που έχουν αναφερθεί στο προηγμένο κεφάλαιο και θα χρησιμοποιηθούν οι μετρικές αυτές για να εξαχθούν διάφορες τάσεις των πελάτων. Στο βήμα της Μοντελοποίησης θα δοκιμαστούν αλγόριθμοι στην κατάτμηση των πελάτων όπως αυτοί αναφερθήκαν στο προηγούμενο κεφάλαιο και θα ερμηνευθούν τα αποτελέσματα της διαδικασίας.

4.1 Business understanding Phase in Crisp-DM

Οπως αναφέρθηκε και νωρίτερα, σκοπός της παρούσας διπλωματικής εργασίας είναι να δείξει πως μπορούν να αξιοποιηθούν νέες τεχνολογίες από συσκευές Διαδικτύου των Πραγμάτων (IoT), για να βοηθήσουν τους λιανέμπορους να κατανοήσουν τη συμπεριφορά του καταναλωτή.

μέσα σε ένα κατάστημα λιανικής. Με χρήση της τεχνολογίας ανάλυσης θέσης (indoor location systems ILS), δίνεται η δυνατότητα στη διοίκηση ενός καταστήματος λιανικής, να κατανοήσει τον τρόπο κίνησης των πελατών μέσα στο κατάστημα, εξάγοντας χρήσιμα συμπεράσματα σχετικά με τα διαφορετικά χαρακτηριστικά της αγοραστικής τους συμπεριφοράς (προϊόντα που προτιμούν, σημεία του καταστήματος στα οποία παραμένουν περισσότερο και επομένως αυξάνουν τις πιθανότητες αγοράς για τα προϊόντα που είναι τοποθετημένα εκεί κλπ.). Στην συνέχεια παρουσιάζονται οι δυνατότητες που προσφέρει η ανάλυση και η εξόρυξη χωροχρονικών δεδομένων, προκειμένου να μπορέσουν οι λιανέμποροι να αναγνωρίσουν και να εντοπίσουν τις διαφορετικές ομάδες των καταναλωτών και τα χαρακτηριστικά αυτών. Με αυτόν τον τρόπο οι λιανέμποροι θα είναι σε θέση να βελτιώσουν τη θετική εμπειρία των καταναλωτών μέσα στο κατάστημα, να παρέχουν καινοτόμες υπηρεσίες στους πελάτες τους και ανταγωνιστικά προϊόντα που να ανταποκρίνονται καλύτερα στις ανάγκες τους, κατά τρόπο παρόμοιο με αυτό που συμβαίνει στα διαδικτυακά καταστήματα (e-shops).

Καθώς η συμπεριφορά των καταναλωτών αλλάζει, επισκεπτόμενοι ένα κατάστημα κάθε φορά για να εκπληρώσουν και μια διαφορετική τους ανάγκη, οι πελάτες γίνονται όλο και πιο απαιτητικοί. Γι' αυτό το λόγο, η υιοθέτηση τεχνολογιών βασισμένων στην Εξόρυξη Γνώσης (Data Mining) και στην Επιχειρηματική Ευφυΐα (Business Intelligence ή BI) από τη διοίκηση ενός καταστήματος λιανικής, τη βοηθάει να ανακαλύψει τις καταναλωτικές τάσεις, προτιμήσεις και συνήθειες των πελατών του, κι έτσι να προβλέψει τις ανάγκες τους, να τους προωθήσει προσωποποιημένες προσφορές προϊόντων και να βελτιώσει τη συνολική εμπειρία τους κατά το χρόνο παραμονής τους στο κατάστημα. Δηλαδή, η χρήση των παραπάνω τεχνολογιών, υποβοηθεί τη διοίκηση του καταστήματος στη λήψη επιχειρηματικών αποφάσεων, βασισμένων σε αξιόπιστα δεδομένα, αυξάνοντας έτσι τις πιθανότητες για τη δημιουργία νέων πωλήσεων και αύξησης της κερδοφορίας της εταιρείας.

Στην παρούσα εργασία εξετάζεται συγκεκριμένα ο τρόπος «ενοποίησης» (omni-channel integration) της πληροφορίας που προέρχεται αφενός από τα συστήματα διαχείρισης του e-shop της εταιρείας (e-commerce, Google Analytics κλπ.) και αφετέρου από ένα σύστημα ανάλυσης θέσης (indoor location system ILS) του φυσικού καταστήματος της, με τη χρήση κατάλληλων μετρικών, οι οποίες δίνουν τη δυνατότητα σύγκρισης της συμπεριφοράς του καταναλωτή τόσο στο e-shop όσο και στο φυσικό κατάστημα. Οι συγκεκριμένες αυτές μετρικές, θα πρέπει να διαθέτουν χαρακτηριστικά, τα οποία θα τις καθιστούν εφαρμόσιμες για την αξιολόγηση της απόδοσης τόσο του e-shop όσο και του φυσικού καταστήματος της εταιρείας. Με αυτό τον τρόπο, η διοίκηση του καταστήματος θα μπορεί να έχει στη διάθεσή της μία ενοποιημένη αναφορά πληροφοριών για να μπορεί να συγκρίνει τα δύο διαφορετικά καταστήματά της (ηλεκτρονικό και φυσικό).



4.2 Δεδομένα - Data Understanding Phase in Crisp-DM

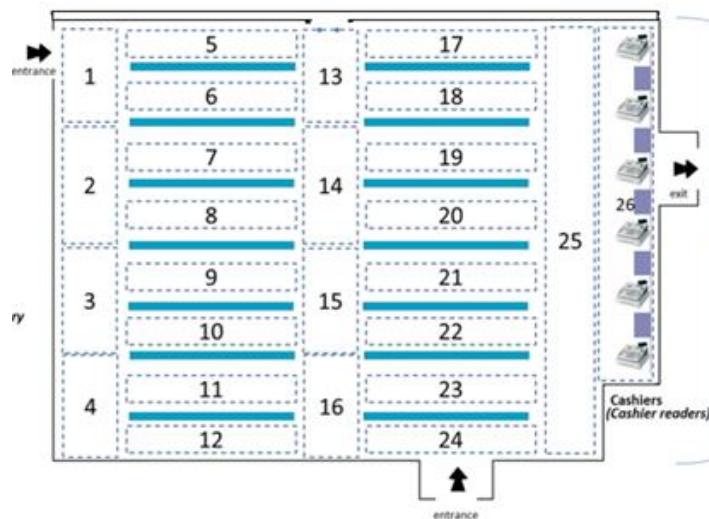
To dataset αφορά δεδομένα, των οποίων η συλλογή έχει γίνει από Bluetooth Low Energy (BLE) beacons. Συγκεκριμένα, τα δεδομένα αυτά αφορούν κινήσεις πελατών μέσα σε ένα κατάστημα λιανικής πώλησης ρούχων. Οι κινήσεις αυτές καταγράφονται καθώς οι πελάτες κινούνται μέσα στο κατάστημα, μέσω μιας mobile εφαρμογής που έχουν εγκατεστημένη στο κινητό τους. Στη συνέχεια τα δεδομένα αυτά αποθηκεύονται σε μια cloud-based βάση δεδομένων.



Εικόνα 8: Επισκόπηση συστήματος ευρέσεως θέσης με βάση την τεχνολογία BLE

4.2.1 Περιγραφή δεδομένων – Describe data Step

Πιο συγκεκριμένα, στον παρακάτω πίνακα 2 μπορούμε να δούμε ένα δείγμα των δεδομένων από το αρχείο dataset. Το «customer_id» είναι μοναδικό για κάθε πελάτη. Κάθε φορά που ένας πελάτης επισκέπτεται το κατάστημα, οι κινήσεις του καταγράφονται μέσω της εφαρμογής και δημιουργείται ένα session (επίσκεψη). Κάθε πελάτης μπορεί να έχει επισκεφτεί το κατάστημα παραπάνω από μια φορές, κατά συνέπεια μπορεί να έχει δημιουργήσει πολλά sessions. Σε κάθε session ένας πελάτης μπορεί να έχει επισκεφτεί πολλές διαφορετικές περιοχές (areas) του καταστήματος. Κάθε περιοχή έχει έναν αναγνωριστικό κωδικό και μπορούμε, κοιτώντας την εικόνα 8, να δούμε ο πελάτης που πραγματικά βρίσκεται μέσα στο κατάστημα.



Εικόνα 9: Κάτοψη του καταστήματος

Το πεδίο timestamp δείχνει τη χρονική στιγμή που εισήλθε ο πελάτης σε μια συγκεκριμένη περιοχή. Το timestamp ορίζεται ως ο αριθμός των δευτερολέπτων που έχουν παρέλθει από τα μεσάνυχτα της 1ης Ιανουαρίου του 1970 (UTC). Και η στήλη seq_id δείχνει με ποια σειρά έγιναν οι καταγραφές των επισκέψεων, δηλαδή τη σειρά με την οποία, ένας πελάτης επισκέφτηκε τις περιοχές μέσα στο κατάστημα.

Για παράδειγμα παρατηρώντας την πρώτη εγγραφή του πίνακα 2 συμπεραίνουμε τα εξής: Ο cust0 - πρώτος εγγεγραμμένος πελάτης στην εφαρμογή για την πρώτη του επίσκεψη στο κατάστημα (ses 0) επισκέφτηκε πρώτη (seq_id= 0) την περιοχή 24 (area=24) που είναι η πλησιέστερη περιοχή σε μια από τις δύο εισόδους, την χρονική στιγμή 1409809816 (timestamp= 1409809816).

	customer_id	session	timestamp	area	seq_id
0	cust0	ses0	1409809816	24	0
1	cust0	ses0	1409809834	25	1
2	cust0	ses0	1409809943	20	2
3	cust0	ses0	1409809974	25	3
4	cust0	ses0	1409810047	21	4

Πίνακας 2: Δεδομένα καταγραφής κίνησης πελατών στο κατάστημα

Επίσης σε ένα δεύτερο dataset περιέχονται οι περιοχές του καταστήματος με την αντιστοίχιστη τους με τις κατηγορίες των προϊόντων της κάθε περιοχής πχ. η area1 περιέχει προϊόντα της κατηγορίας cloak.

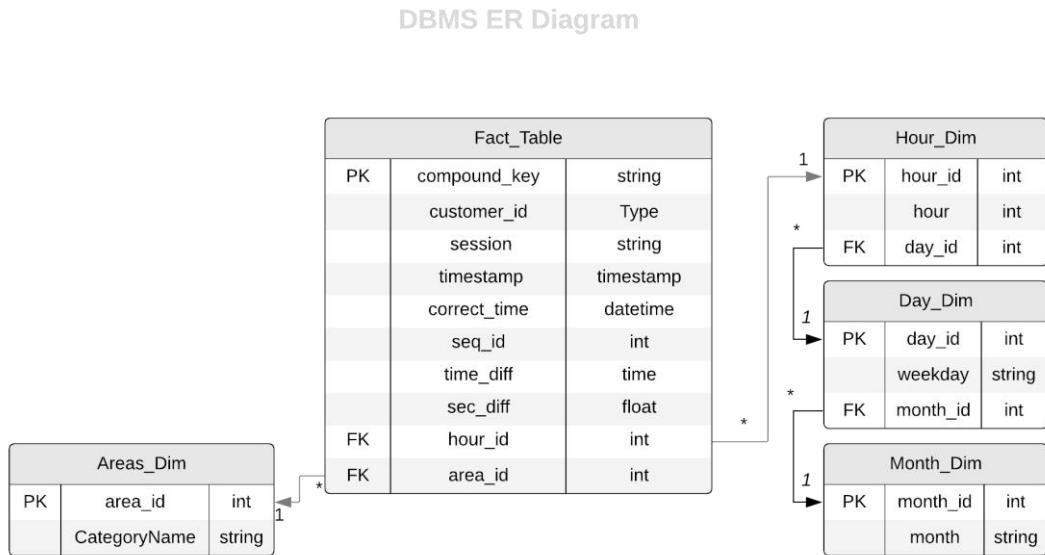
CategoryID	CategoryName
0	1 cloak
1	2 overcoat+raincoat
2	3 cloak+1accessory
3	4 jacket
4	5 jacket+skirt

Πίνακας 3: Αντιστοιχία περιοχών καταστήματος με το αναγνωριστικό κωδικό της

Το δεύτερο dataset συνδυάστηκε με την Εικόνα 9 για να παραχθεί η Εικόνα 7, όπου απεικονίζεται η χωρική θέση των κατηγοριών των προϊόντων μέσα στο κατάστημα. Αυτή η απεικόνιση του χώρου του καταστήματος χρησιμοποιήθηκε για την οπτικοποίηση των δεικτών απόδοσης (KPIs, βλέπε κεφάλαιο 5).

4.3 Προετοιμασία Δεδομένων -Data Preparation Phase in Crisp-DM

Στην παρακάτω εικόνα απεικονίζεται το διάγραμμα οντοτήτων-συσχετίσεων (ERD) του σχεσιακού μοντέλου που χρησιμοποιήθηκε στην εργασία.



Εικόνα 10: Διάγραμμα οντοτήτων-συσχετίσεων (ERD)

4.3.1 Δημιουργία νέων πεδίων δεδομένων – Construct data Step

Το πεδίο **timestamp** έπρεπε να μετασχηματιστεί σε αναγνώσιμη μορφή η οποία ονομάστηκε **correct_time**. Στη βάση δεδομένων δημιουργήθηκαν κάποιες οντότητες από το πεδίο **timestamp** όπως η ώρα προσέλευσης (hour), η ημέρα (weekday) και ο μήνας (month) και αυτά αποθηκεύτηκαν στη βάση με ιεραρχική δομή, σχήματος χιονονιφάδας (snowflake schema, βλέπε εικόνα 9). Τα ανωτέρω περιγράφονται αναλυτικά και στον κατωτέρω πίνακα 4, η στήλη **correct_time** αφορά στην ημερομηνία και στην ώρα, η στήλη **time_diff** αφορά στο χρόνο που έμεινε ο πελάτης στην περιοχή, εκφρασμένη σε δευτερόλεπτα. Επίσης για να μελετηθούν τα δεδομένα σε επίπεδο επίσκεψης, κατασκευάστηκε το πεδίο **compound_key**, από το συνδυασμό των πεδίων **customer_id** και **session**. Από μόνα τους δεν αποτελούν μοναδικά κλειδιά αλλά ο συνδυασμός τους είναι μοναδικός π.χ δεν υπάρχει άλλη επίσκεψη **cust0ses0** ενώ υπάρχουν πολλές **cust0** ή **ses0**.

	customer_id	session	compound_key	timestamp	correct_time	area	seq_id	time_diff	CategoryName	weekday	month	hour	sec_diff
0	cust0	ses0	cust0 ses0	1409809816	2014-09-04 08:50:16	24	0	00:00:18	dungaree	Thursday	September	8	18
1	cust0	ses0	cust0 ses0	1409809834	2014-09-04 08:50:34	25	1	00:01:49	accessories	Thursday	September	8	109
2	cust0	ses0	cust0 ses0	1409809943	2014-09-04 08:52:23	20	2	00:00:31	mid-lengthjacket+skirt	Thursday	September	8	31
3	cust0	ses0	cust0 ses0	1409809974	2014-09-04 08:52:54	25	3	00:01:13	accessories	Thursday	September	8	73
4	cust0	ses0	cust0 ses0	1409810047	2014-09-04 08:54:07	21	4	00:00:32	mid-lengthjacket+trousers	Thursday	September	8	32
5	cust0	ses0	cust0 ses0	1409810079	2014-09-04 08:54:39	15	5	00:01:27	gilet	Thursday	September	8	87

Πίνακας 4: Αναθεωρημένος πίνακας με τα νέα πεδία

4.3.2 Συγκεντρωτικά Χαρακτηριστικά -Dataset Description Step

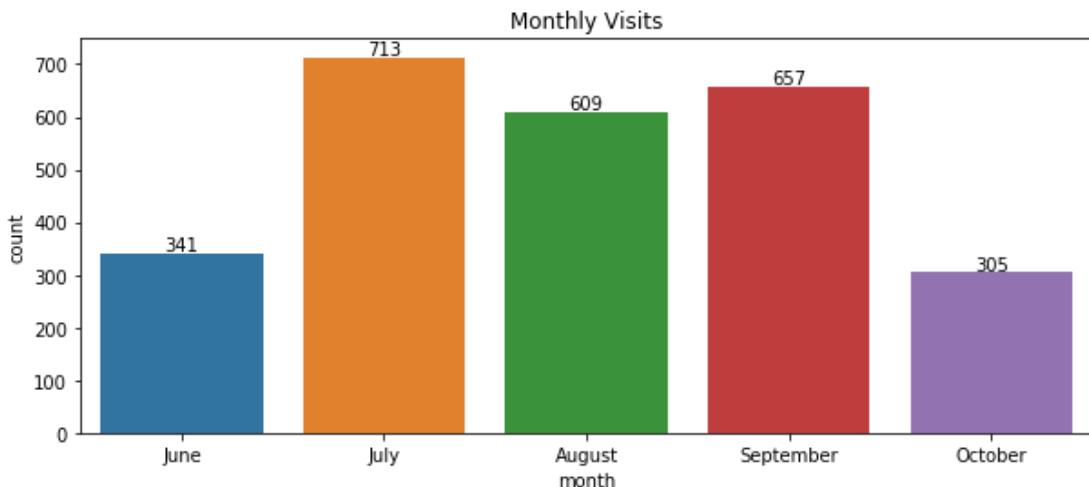
Κάποια στατιστικά που αφορούν τα πεδία μπορούν να παρατηρηθούν στον παρακάτω πίνακα. Υπήρχαν συνολικά 91674 εγγραφές από 600 συνολικά μοναδικούς χρήστες (customer_id unique value) Ο χρήστης με τις περισσότερες εγγραφές ήταν ο cust369. Η επίσκεψη με τις περισσότερες εγγραφές ήταν η έκτη επίσκεψη του χρήστη cust451. Για την σειρά με την οποία έγιναν οι επισκέψεις στις περιοχές (seq_id) παρατηρούνται τα εξής: κατά μέσο όρο στις επισκέψεις τους οι χρήστες επισκέφτηκαν 24 περιοχές και με τυπική απόκλιση 20. Η μεγαλύτερη επίσκεψη είχε 110 εγγραφές περιοχών (πολλαπλές μετακινήσεις στις περιοχές του καταστήματος σε μια επίσκεψη). Τα δοθέντα δεδομενα προς ανάλυση είναι από τις 5/6/2014 23:16:19 εως τις 21/10/2014 22:44:26.

	customer_id	session	compound_key	area	seq_id	CategoryName	weekday	month	hour	sec_diff
count	91674	91674	91674	91,674	91,674	91674	91674	91674	91,674	91,674
unique	600	8	2650	nan	nan	26	7	5	nan	nan
top	cust369	ses0	cust451 ses6	nan	nan	accessories	Monday	July	nan	nan
freq	446	21130	111	nan	nan	10981	16655	24428	nan	nan
mean	NaN	NaN	NaN	14	24	NaN	NaN	NaN	11	63
std	NaN	NaN	NaN	8	20	NaN	NaN	NaN	7	33
min	NaN	NaN	NaN	1	0	NaN	NaN	NaN	0	0
25%	NaN	NaN	NaN	6	8	NaN	NaN	NaN	6	34
50%	NaN	NaN	NaN	14	19	NaN	NaN	NaN	11	63
75%	NaN	NaN	NaN	21	36	NaN	NaN	NaN	17	92
max	NaN	NaN	NaN	26	110	NaN	NaN	NaN	23	120

Πίνακας 5: Πίνακας με στατιστικά των πεδίων

Οπτικοποιήσεις Των Νέων Δεδομένων Explore Data Step

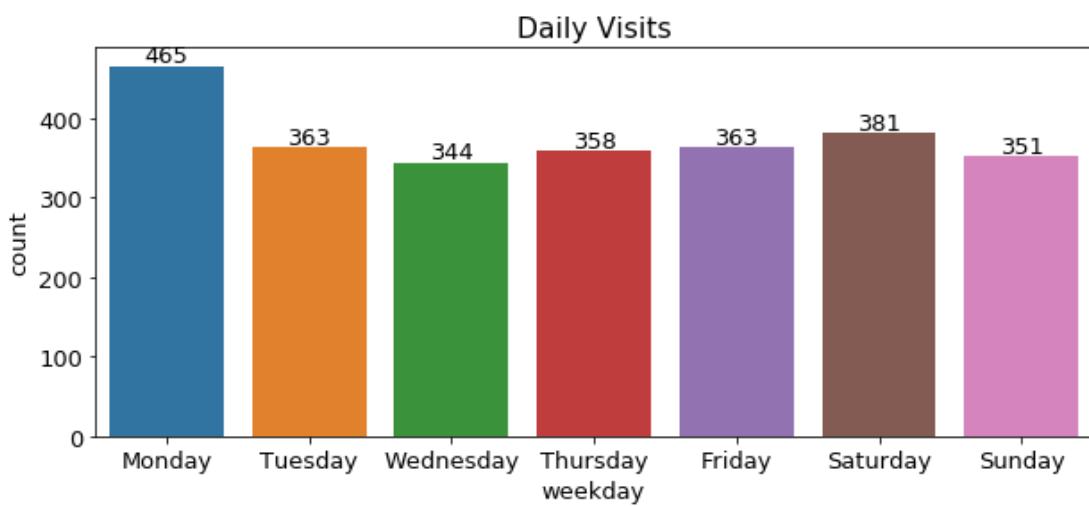
Με τις ανάλογες οπτικοποίησεις μπορούμε να δούμε και άλλα δεδομένα για τη χρονική εξέλιξη των επισκέψεων των πελάτων στο κατάστημα που χρησιμοποίησαν την εφαρμογή.



Εικόνα 11: Αριθμός επισκέψεων ανά Μηνά

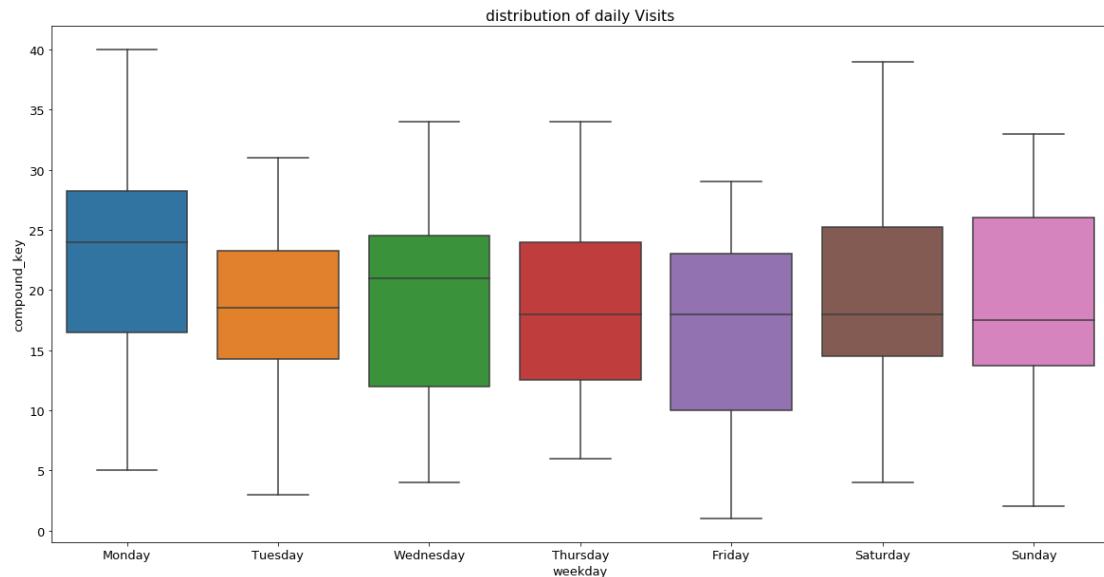
Στην παραπάνω εικόνα παρατηρείται ότι το κατάστημα είχε την μεγαλύτερη επισκεψιμότητα το μήνα Ιούλιο και τους μήνες Ιούνιο και Οκτώβριο, είχε την μικρότερη.

Αντίστοιχα σε επίπεδο εβδομάδας, στην παρακάτω εικόνα φαίνεται ότι το κατάστημα είχε τη μεγαλύτερη επισκεψιμότητα, την ημέρα Δευτέρα με 465 επισκέψεις και τη μικρότερη την ημέρα Τέταρτη με 344.



Εικόνα 12: Αριθμός επισκέψεων ανά Ημέρα

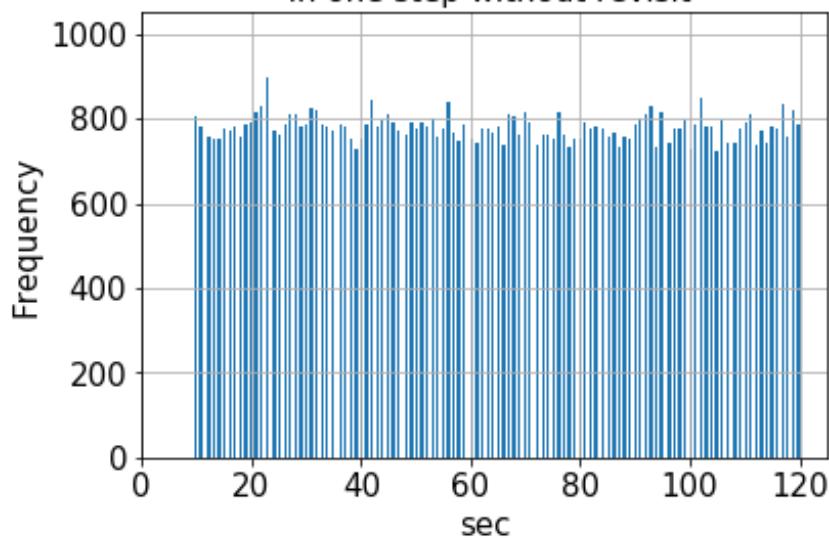
Άλλη σημαντική οπτικοποίηση είναι το θηκόγραμμα(boxplots) που απεικονίζει περισσότερες πληροφορίες, τον διάμεσο μαζί με το 1ο και το 3ο τεταρτημόριο των τιμών.



Εικόνα 13: Θηκόγραμμα των επισκέψεων ανά ημέρα της εβδομάδας

Παρατηρείται ότι οι διάμεσοι της κάθε μέρας δεν διαφέρουν και πολύ πέραν από της Δευτέρας που έχει τις περισσότερες επισκέψεις. Επίσης το ενδοτεταρτημοριακό εύρος (ήτοι το μέγεθος των κουτιών) στα παραπάνω θηκογράμματα είναι μεγαλύτερο για την Παρασκευή, την Κυριακή και την Τέταρτη, δηλαδή αυτές τις ήμερες παρουσιάζεται μεγαλύτερη διακύμανση στον αριθμό των επισκέψεων.

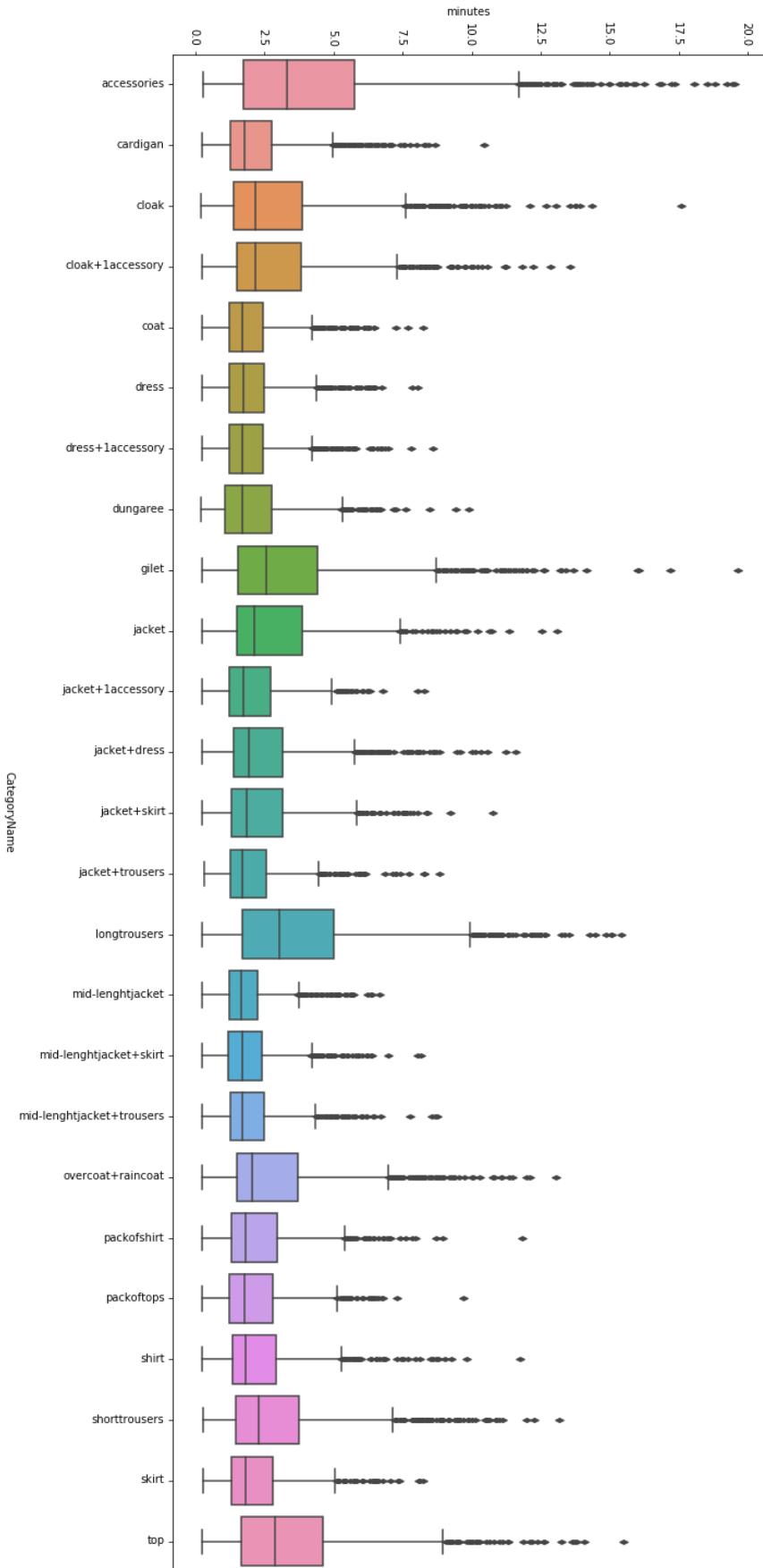
Distribution of seconds for all categories customers visited in one step without revisit



Εικόνα 14: Κατανομή της χρονικής διάρκειας όλων των επισκέψεων σε όλες τις κατηγορίες

Παρατηρείται ότι η διάρκεια επίσκεψης στις περιοχές είναι από 10 έως 120 δευτερόλεπτα και είναι σχεδόν ισοπίθανες.

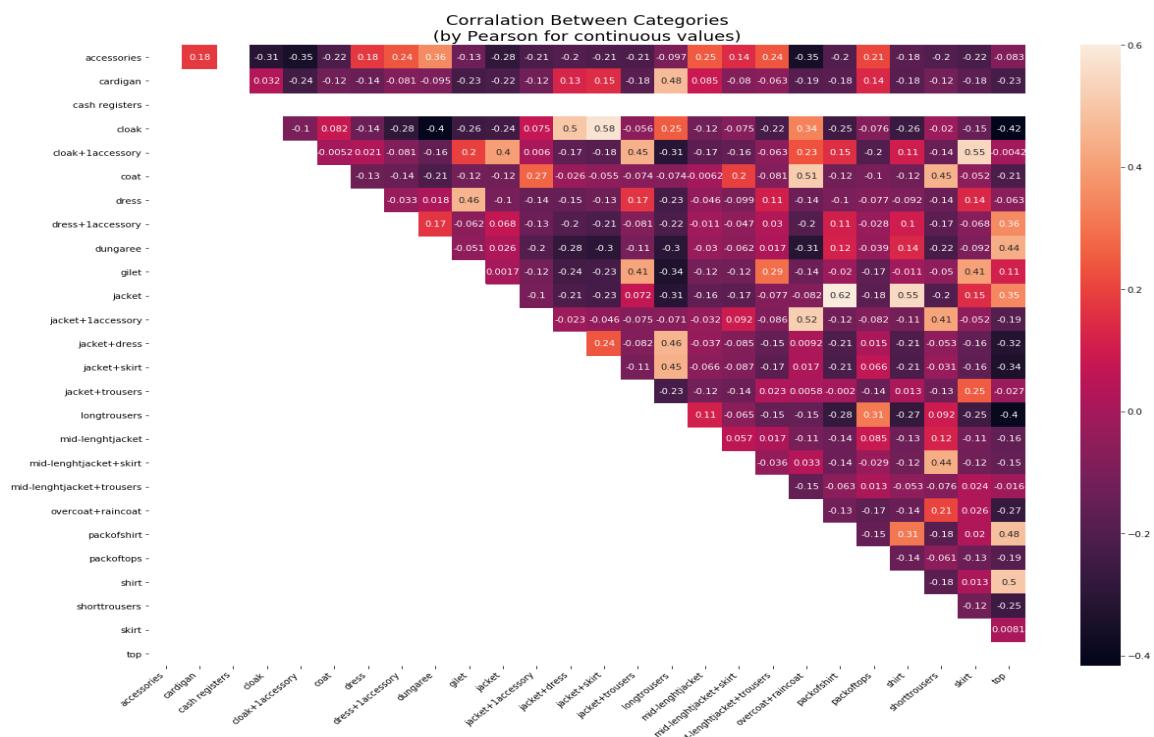
Στην παρακάτω εικόνα απεικονίζονται συγκεντρωτικά όλες οι περιοχές (κάθετα) με τις διάρκειες επίσκεψης τους σε θηλογράμματα (οριζόντια). Όλες οι περιοχές παρουσιάζουν θετική ασυμετρία (right-skewed) δηλαδή όλες οι περιοχές είχαν επισκέψεις πλειοψηφικά μικρής διάρκειας και ταυτόχρονα είχαν κάποιες μακράς διάρκειας επισκέψεις αλλιώς ακραίες τιμές (outliers). Γι' αυτό το λόγο, παρακάτω στη διπλωματική αποφασίστηκε να γίνει ένας μετασχηματισμός των πεδίων και να μετατραπούν οι τιμές από αριθμητικές σε δυαδικές (επισκέφτηκε=1, δεν επισκέφτηκε =0).



Εικόνα 15: violin plot των χρόνων επίσκεψης ανά κατηγορία

4.3.3 Επιλογή Δεδομένων - Select Data Step

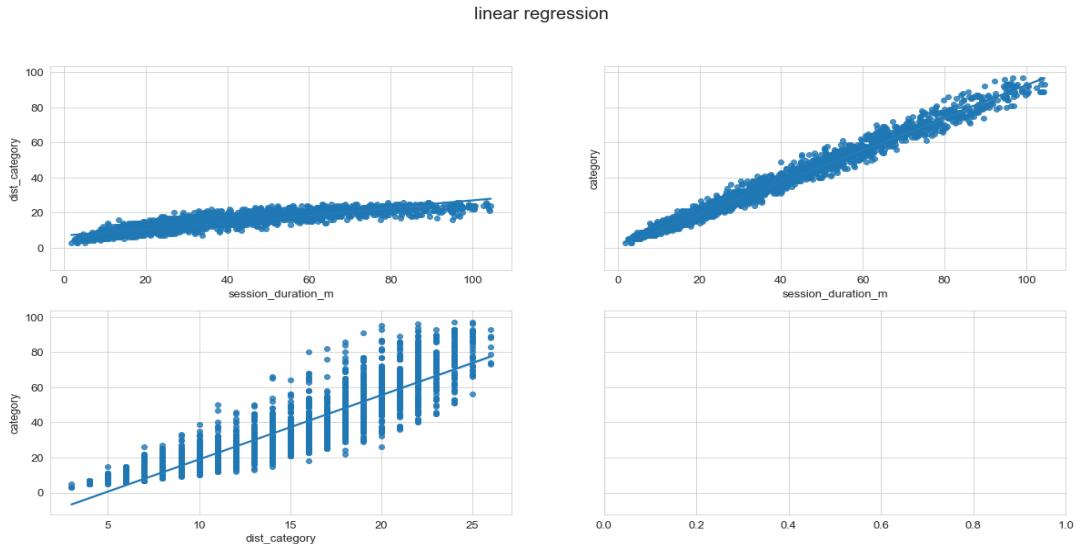
Όταν γίνεται μελέτη ενός αριθμού δεδομένων με δυο και περισσότερα χαρακτηριστικά ένα πρώτο βήμα πριν την ανάλυση είναι να εκφραστεί και να μελετηθεί η πιθανή συσχέτιση των χαρακτηριστικών. Για παράδειγμα, στην περίπτωση του χαρακτηριστικού του χρόνου παραμονής των πελατών σε μια κατηγορία προϊόντων, θα μελετηθεί κατά πόσο ισχύει ότι όσο περισσότερο χρόνο περνάνε οι πελάτες σε μια κατηγορία προϊόντων τόσο περισσότερο χρόνο περνάνε και σε μια άλλη. Για αυτό το σκοπό χρησιμοποιείται ο Συντελεστής Γραμμικής Συσχέτισης του Pearson που συμβολίζεται με ρ . Στην παρακάτω εικόνα παρουσιάζονται οι κατηγορίες με τη μεταξύ τους συσχέτιση σε ένα heatmap, όπου δεξιά είναι η στήλη με την κλίμακα του ρ . Έτσι μπορεί να βρεθεί σε ποιες δυο περιοχές/κατηγορίες προϊόντων τείνουν να περνούν οι πελάτες πολύ χρόνο σε μια επίσκεψη για θετικό ρ και το αντίθετο για αρνητικό ρ^5 . Αφού βρεθούν οι συσχέτισεις πρέπει να διερευνηθεί η αιτιότητα αυτών καθώς μπορεί να βρεθούν σχέσεις μεταξύ των κατηγοριών π.χ. για παράλληλη πώληση (cross sale) για θετικό ρ και για αρνητικό ρ προϊόντα που δεν συνδυάζονται μεταξύ τους και δεν έχει νόημα να είναι μαζί σε πρωθητικές ενέργειες. Από την ανάλυση προέκυψε ότι δεν υπάρχουν κατηγορίες με υψηλή συσχέτιση που μπορούν να επηρεάσουν την μετέπειτα μοντελοποίηση οπότε δεν απορρίπτεται κάποια κατηγορία.



Εικόνα 16: Συσχέτιση μεταξύ των κατηγοριών με μετρική τον χρόνο των πελατών ανά κατηγορία

⁵.Το ρ παίρνει τιμές στο κλειστό διάστημα [-1, 1]. Για τιμές του ρ έχουμε την σύμβαση : αν $\rho > 0,7$ υπάρχει ισχυρή θετική συσχέτιση, από 0,5 έως 0,7 θετική συσχέτιση και ασθενή από 0,3 έως 0,5. Ενώ για μικρότερες δεν υπάρχει

Επίσης σημαντικό είναι να μελετηθεί η σχέση μεταξύ της διάρκειας των επισκέψεων και του αριθμού των κατηγοριών.



Πίνακας 6: Διαγράμματα διασποράς και ευθείας ελαχίστων τετραγώνων για το χρόνο επίσκεψης και τον αριθμό των περιοχών επίσκεψης και μοναδικών περιοχών επίσκεψης

OLS Regression Results						
Intercept	1.497837	session_duration_m	0.911542			
dtype: float64						
Dep. Variable:	category	R-squared:	0.983			
Model:	OLS	Adj. R-squared:	0.983			
Method:	Least Squares	F-statistic:	1.506e+05			
Date:	Tue, 27 Feb 2020	Prob (F-statistic):	0.00			
Time:	13:34:11	Log-Likelihood:	-6469.4			
No. Observations:	2625	AIC:	1.294e+04			
Df Residuals:	2623	BIC:	1.295e+04			
Df Model:	1					
Covariance Type:	nonrobust					
=====						
coef	std err	t	P> t	[0.025	0.975]	
Intercept	1.4978	0.100	14.918	0.000	1.301	1.695
session_duration	0.9115	0.002	388.132	0.000	0.907	0.916
=====						
Omnibus:	87.304	Durbin-Watson:		1.974		
Prob(Omnibus):	0.000	Jarque-Bera (JB):		138.461		
Skew:	0.303	Prob(JB):		8.58e-31		
Kurtosis:	3.948	Cond. No.		77.3		
=====						

Από την ανωτέρω γραφική απεικόνιση (διάγραμμα διασποράς) παρατηρήθηκε ότι υπάρχει μια γραμμική σχέση ανάμεσα στον αριθμό των κατηγοριών/ περιοχών που επισκέφθηκαν οι πελάτες με το χρόνο παραμονής στο κατάστημα.

Στη συνέχεια αυτού του συμπεράσματος μελετήθηκε περαιτέρω η εν λόγω σχέση μέσω της γραμμικής παλινδρόμησης και παρατηρήθηκε ότι αν ένας πελάτης επισκεφθεί μια ακόμα περιοχή θα αυξηθεί ο χρόνος παραμονής του στο κατάστημα κατά 0,91 λεπτά (ή 55 δευτερόλεπτα).

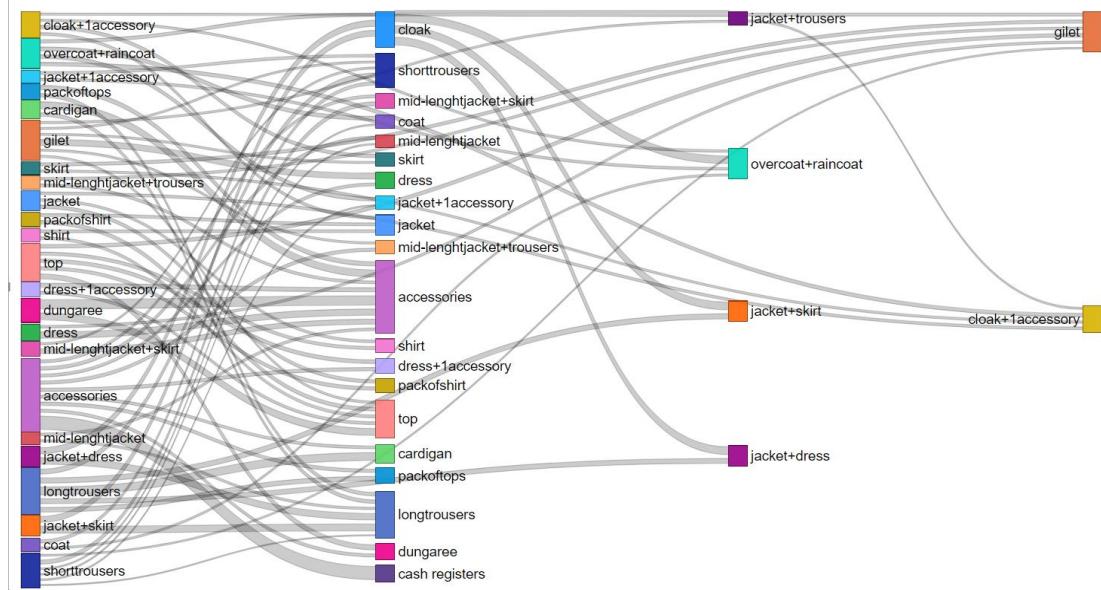
$$\text{Category} = 0.9115 * \text{session_duration_m} + 1.497837 + \epsilon$$

με R-squared: 0.983

4.3.4 Οπτικοποίηση Των Μετακινήσεων Των Πελάτων Μέσω Γράφων Explore

Data Step

Για να βρεθούν σχέσεις μεταξύ των κατηγοριών που επισκέφτηκε οι πελάτες μπορεί να χρησιμοποιηθεί το διάγραμμα Sankey, όπως αυτό φαίνεται στην παρακάτω εικόνα.

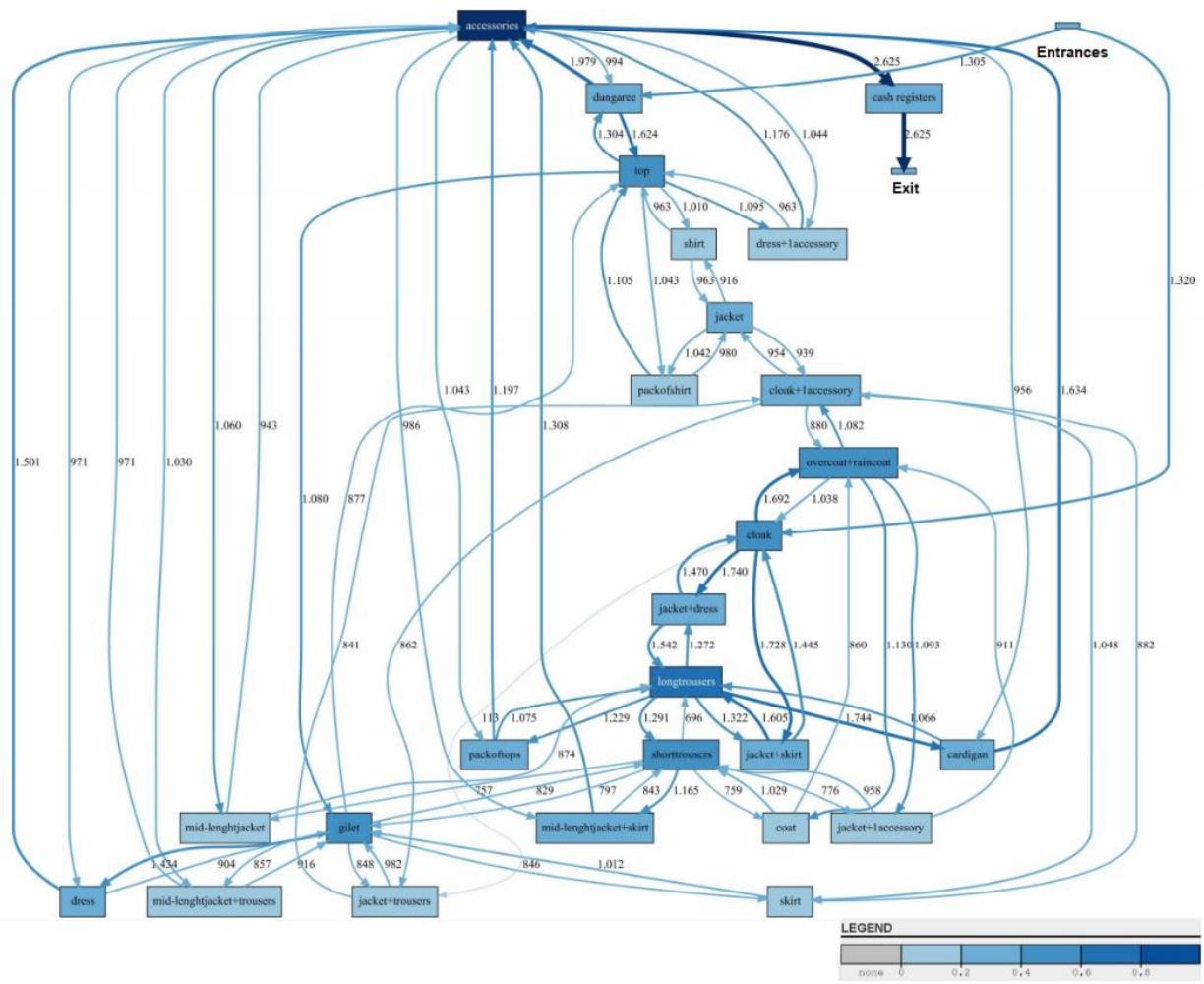


Εικόνα 17: Sankey chart με τις μετακινήσεις των πελάτων στο κατάστημα

Παρατηρώντας το διάγραμμα, είναι εύκολο να εντοπιστούν οι μεταβάσεις από μια κατηγορία σε μια άλλη. Η κάθε κατηγορία έχει διαφορετικό χρώμα και το μήκος της μπάρας δείχνει ποιοτικά τον αριθμό των επισκέψεων της εκάστοτε κατηγορία. Αυτό βοηθάει τον αναλυτή στο να βγάλει γρήγορα συμπεράσματα και να δει τις σχέσεις που έχουν οι κατηγορίες. Πχ η κατηγορία accessories, που έχει τις περισσότερες επισκέψεις, παρατηρείται ότι έχει και τις περισσότερες σχέσεις. Οι πελάτες μετέβησαν σε αυτήν την κατηγορία από 8 διαφορετικές κατηγορίες/περιοχές του καταστήματος και από την κατηγορία accessories επέστρεψαν στις ιδίες 8 κατηγορίες συν την περιοχή των ταμιών.

Για μια πιο ενδελεχή ανάλυση των σχέσεων των κατηγοριών χρειάζεται και η ποσοτική ανάλυσή τους που ακολουθεί.

Συγκεκριμένα οι εγγραφές που χρησιμοποιήθηκαν ήταν 89024 μετακινήσεις των καταναλωτών μέσα στο κατάστημα από μια κατηγορία προϊόντων (αρχική θέση) στην ακριβώς επόμενη κατηγορία προϊόντων (τελική θέση) που επισκέφτηκαν. Στην παρακάτω εικόνα παρουσιάζεται το άθροισμα των μετακινήσεων από κατηγορία σε κατηγορία.

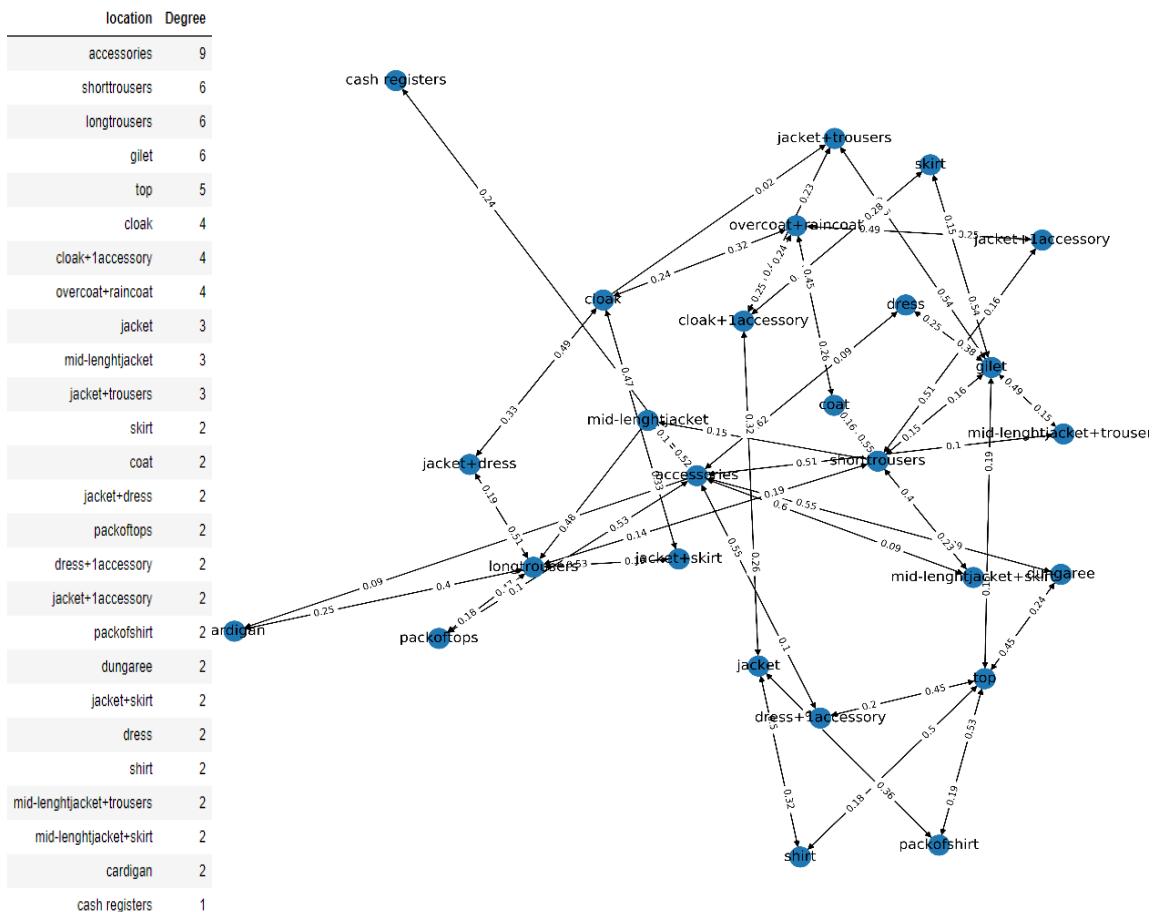


Εικόνα 18: Γράφος με τις μετακινήσεις των καταναλωτών από κατηγορία σε κατηγορία.

Η παραπάνω εικόνα αποτελεί έναν κατευθυντικό γράφο⁶ με κόμβους τις κατηγορίες προϊόντων. Οι κατηγορίες προϊόντων στις οποίες οι καταναλωτές επέλεξαν να μετακινηθούν μεταξύ τους, ενώνονται με κατευθυνόμενες ακμές (βέλη) όπου η ακμή του βέλους δείχνει την κατηγορία στην οποία μετακινήθηκε ο καταναλωτής και η ουρά του βέλους την κατηγορία από την οποία προήλθε. Τα δε βάρη των ακμών δείχνουν το πλήθος των συγκεκριμένων μετακινήσεων μεταξύ των κατηγοριών. Το βάθος χρώματος του κόμβου είναι ανάλογο με το εύρος του αριθμού των πελατών που καταλήγουν σ' αυτόν. Συγκεκριμένα οι πιο σκούρες απεικονίσεις αντιστοιχούν σε μεγαλύτερο πλήθος επισκέψεων, η δε σχετική διαβάθμιση επεξηγείται στη λεζάντα (legend). Πάνω δεξιά της εικόνας διακρίνονται οι είσοδοι του καταστήματος και μπορεί να διαπιστωθεί ότι οι πελάτες χρησιμοποιούν και τις δύο εισόδους του καταστήματος ισοπίθανα. Από την είσοδο, κοντά στην κατηγορία dungaree οι πελάτες εισήλθαν 1305 (49,7%) φορές και από την

⁶ https://el.wikipedia.org/wiki/Κατευθυνόμενος_γράφος

είσοδο, κοντά στην κατηγορία cloak εισήλθαν 1320 (50,3%) φορές, στις συνολικά 2625 επισκέψεις τους. Από το διάγραμμα προκύπτει ότι η κίνηση των πελατών στο κατάστημα δεν είναι προκαθορισμένη.



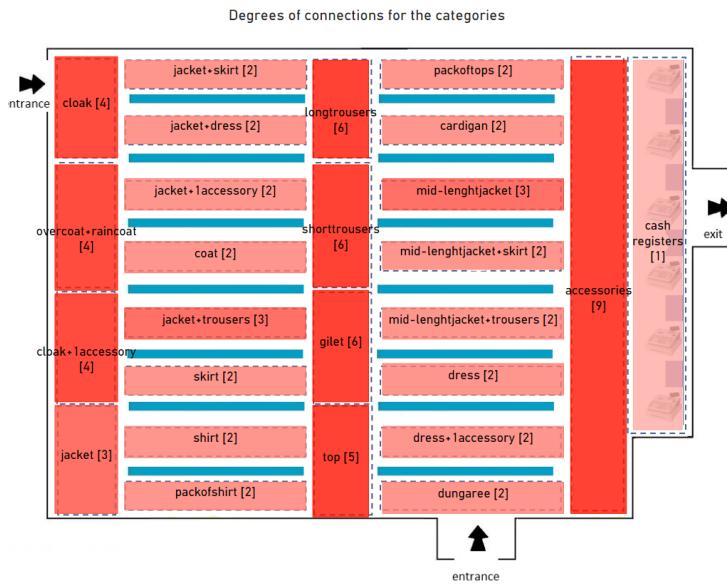
Εικόνα 19: Αριστερά (πίνακας) οι κατηγορίες και ο βαθμός των ακμών που τους αντιστοιχούν. Δεξιά ο Γράφος που αναπαριστά τις μετακινήσεις ανάμεσα στις κατηγορίες εκφρασμένες σε ποσοστά (εικόνα)⁷

Στην παραπάνω εικόνα (αριστερά) φαίνεται ο αριθμός των συνδέσεων (αριθμός των ακμών που καταλήγουν σε μια κατηγορία) και (δεξιά) φαίνονται οι κατηγορίες των προϊόντων με τα ποσοστά μετάβασης σ' αυτά.

Η οπτικοποίηση των αριθμών των συνδέσεων φαίνεται στην επόμενη εικόνα. Είναι ξεκάθαρο ότι το κατάστημα χρησιμοποιεί κλειστούς διαδρόμους, με τους πελάτες να μετακινούνται στους οριζόντιους διαδρόμους μόνο μεταξύ των άμεσα γειτονικών τους περιοχών. Με εξαίρεση την περιοχή με τα mid-length jacket όπου οι πελάτες μεταβαίνουν από τις άμεσα γειτονικές περιοχές με τα short trousers και accessories αλλά και από την περιοχή με τα long trousers που

⁷ Στο τέλος της εργασία παρατίθεται στον αναγνώστη ο πίνακας με τις κατηγορίες και τα ποσοστά και οι απόλυτες τιμές, όπου τα νούμερα να είναι πιο ευκρινή

δεν είναι άμεσα γειτονική. Αυτό συμβαίνει, χωρίς να ισχύει η αντίστροφη σχέση, δηλαδή ο πελάτης να μετακινείται από τα long trousers, προς τα mid-length jacket. Παρόμοια, για την περιοχή jacket+trousers που έχει 3 σχέσεις.



Εικόνα 20: Αριθμός συνδέσεων ανά κατηγοριών

Μ' αυτού του είδους τη γραφική αναπαράσταση των μετακινήσεων ανάμεσα στις κατηγορίες των προϊόντων μπορούν να εξαχθούν διάφορα συμπεράσματα που δύναται να χρησιμοποιηθούν ανάλογα με τις επιχειρησιακές ανάγκες του καταστήματος.

4.4 Μοντελοποίηση Modeling Phase in Crisp-DM

Εδώ θα παρουσιαστεί το κύριο κομμάτι της διπλωματικής, που είναι η χρήση κατάλληλων μεθόδων και αλγορίθμων για την επίλυση του προβλήματος που έχει οριστεί. Στη συνέχεια γίνεται παρουσίαση των αποτελεσμάτων της ανάλυσης των δεδομένων.

4.4.1 Τροποποίηση Δεδομένων Πριν Την Εισαγωγή Τους Στο Μοντέλο Construct

Data And Transform Data Step

Για να υλοποιηθεί η συσταδοποίηση και να συγκριθούν οι αγοραστικές αποστολές, ήταν αναγκαίο να τροποποιηθούν κατάλληλα τα δεδομένα. Για αυτό το λόγο, τα δεδομένα ομαδοποιήθηκαν σε επισκέψεις και κατηγορίες. Τα δεδομένα πήραν δυαδικές τιμές. Από τα δεδομένα των επισκέψεων προκύπτει ποιες κατηγορίες προϊόντων έχουν κοινή προτίμηση. Δηλαδή ποιες κατηγορίες τείνουν να εμφανίζονται μαζί σε μια επίσκεψη.

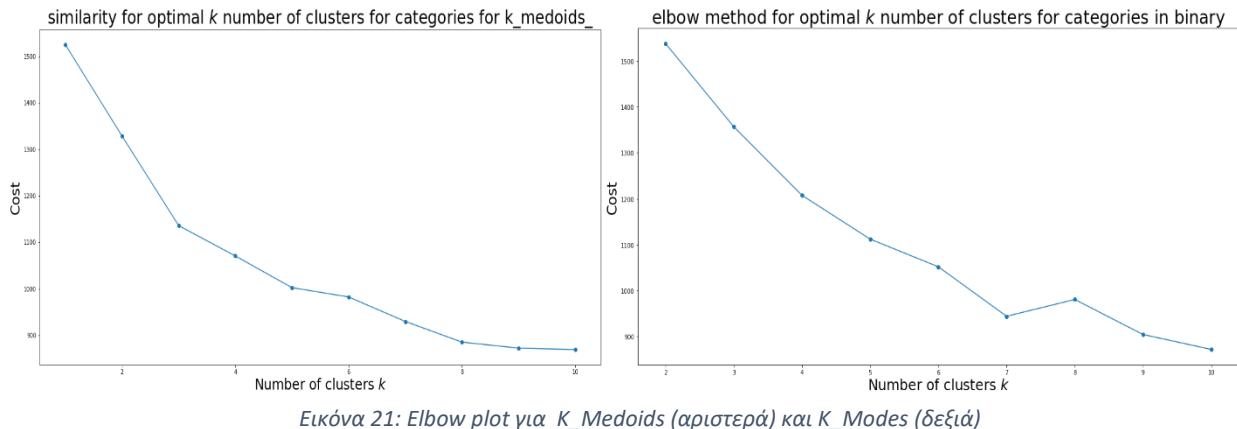
Η απόσταση **Jaccard** έχει την ιδιότητα να μετρά την ομοιότητα δύο αντικειμένων που αποτελούνται από ζευγάρια καταστάσεων οι οποίες δεν είναι ισότιμες μεταξύ τους. Οι καταστάσεις παίρνουν μόνο δύο τιμές 0 και 1 και συμβολίζουν διαφορετικής αξίας πληροφορία.

Στην παρούσα εργασία, δίνεται η τιμή: 0, στην περίπτωση που ο πελάτης δεν επισκέφτηκε μια περιοχή του καταστήματος όπου βρίσκεται συγκεκριμένη κατηγορία προϊόντων και δίνεται η τιμή: 1, στην περίπτωση που ο πελάτης επισκέφτηκε μια περιοχή του καταστήματος όπου βρίσκεται συγκεκριμένη κατηγορία προϊόντων. Η μη επίσκεψη σε μια κατηγορία (τιμή 0) δεν έχει το ίδιο ενδιαφέρον με την επίσκεψη σε μια κατηγορία (τιμή 1). Συνεπώς το γεγονός ότι επισκέφτηκαν δυο χρήστες (παρατηρήσεις) την ίδια κατηγορία προϊόντων (1,1) είναι πιο σημαντικό από το αντίθετο (0,0). Επιπλέον, είναι ασαφής ο λόγος της κοινής μη επίσκεψης (0,0). Για τους παραπάνω λόγους επιλέχθηκε να μην συνυπολογίζεται αυτό το ζευγάρι τιμών στη μέτρηση απόστασης. Κατά συνέπεια, επιλέχθηκε η απόσταση **Jaccard** γιατί μπορεί να απεικονίζει την ασυμμετρία των μεταβλητών «αγνοώντας» ζευγάρια τιμών (παρατηρήσεις) που δεν είναι χρήσιμα για την εξαγωγή συμπερασμάτων.

4.4.2 Σύγκριση Αλγορίθμων Για Την Ανάλυση Των Δεδομένων Assess Mode Step

Elbow Plot

Αυτή η μέθοδος, όπως έχει αναφερθεί στη θεωρία, προσπαθεί να βρει το κατάλληλο αριθμό συστάδων που εξηγεί καλύτερα τα δεδομένα. Για να εφαρμοστεί η μέθοδος ο κάθε αλγόριθμος τρέχει για ένα εύρος αριθμού συστάδων, το εύρος που επιλέχτηκε στην εργασία ήταν το [2,10].



Παρατηρείται ότι και στους δυο τύπους διαιρετικών αλγόριθμων K-Medoids και K-Modes δεν υπάρχει κάποιος «χαρακτηριστικός» αριθμός (σημείο στη γραφική παράσταση) όπου αλλάζει η κλίση της καμπύλης. Για τον K - Medoids και για τον K - Modes ένα τέτοιο σημείο από τις γραφικές είναι ο αριθμός 5, για τον K - Modes επιπλέον, είναι και αριθμός 7.

Silhouette Score

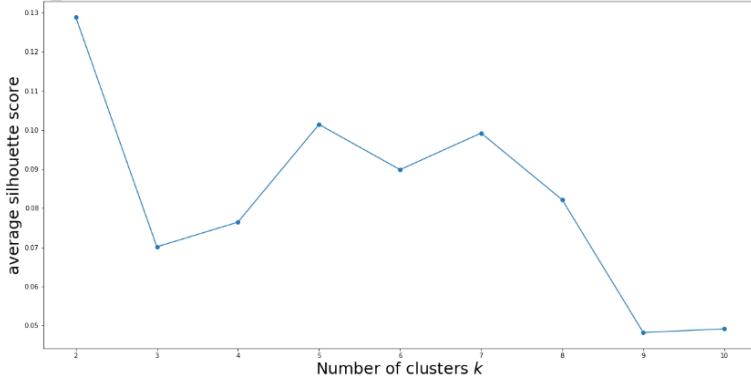
Η μέθοδος αυτή καθορίζει τον αριθμό των συστάδων με βάση την ομοιότητα του κάθε στοιχείου με τα υπόλοιπα στοιχεία της συστάδας στην οποία ανήκει και συγκρίνοντας το με τα δεδομένα των υπόλοιπων συστάδων. Με αυτό τον τρόπο βρίσκει κατά πόσο μια παρατήρηση είναι σωστά τοποθετημένη στη συστάδα που την έχει «αναθέσει» ο αλγόριθμος. Η τιμή του Silhouette score κυμαίνεται από -1 έως 1, όπου όσο υψηλότερη είναι η τιμή τόσο δηλώνει ότι ένα στοιχείο είναι καλά τοποθετημένο στο συγκεκριμένο cluster και ταιριάζει ελάχιστα με τα γειτονικά. Για να εφαρμοστεί η μέθοδος, ο κάθε αλγόριθμος τρέχει για ένα εύρος αριθμού συστάδων, το εύρος που επιλέχτηκε στην εργασία ήταν το [2,10].

```

For n_clusters = 2 The average silhouette_score is : 0.12886957221254972
For n_clusters = 3 The average silhouette score is : 0.07012957932478214
For n_clusters = 4 The average silhouette score is : 0.07642662881127843
For n_clusters = 5 The average silhouette score is : 0.10141227208063507
For n_clusters = 6 The average silhouette score is : 0.08985264114338935
For n_clusters = 7 The average silhouette score is : 0.09920567707114751
For n_clusters = 8 The average silhouette score is : 0.08216014644940833
For n_clusters = 9 The average silhouette score is : 0.048211267977339184
For n_clusters = 10 The average silhouette score is : 0.049164957899120496

```

silhouette_score for optimal k number of clusters for categories binary with jaccard distance



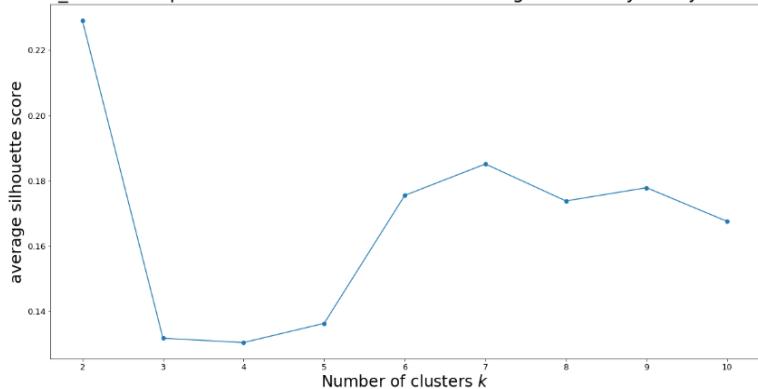
Εικόνα 22: Αποτελέσματα της Silhouette score για K-medoids

```

For n_clusters = 2 The average silhouette_score is : 0.22908544130050573
For n_clusters = 3 The average silhouette_score is : 0.13172502933898725
For n_clusters = 4 The average silhouette_score is : 0.13039467361539359
For n_clusters = 5 The average silhouette_score is : 0.13627619641392336
For n_clusters = 6 The average silhouette_score is : 0.17548301917438705
For n_clusters = 7 The average silhouette_score is : 0.18510039990478808
For n_clusters = 8 The average silhouette_score is : 0.17378195266958923
For n_clusters = 9 The average silhouette_score is : 0.17783937487805745
For n_clusters = 10 The average silhouette_score is : 0.16751425566043174

```

silhouette_score for optimal k number of clusters for categories binary with jaccard distance



Εικόνα 23: Αποτελέσματα της Silhouette Score για του K modes

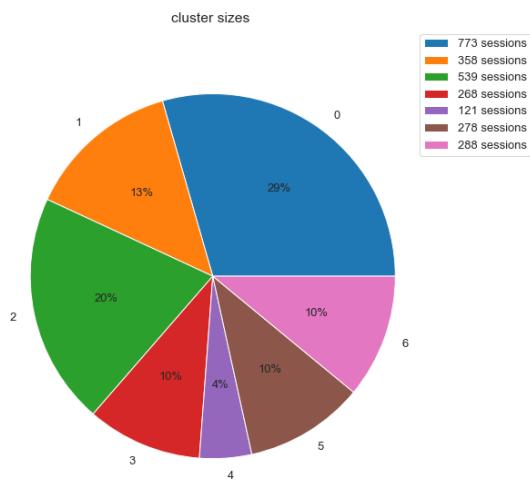
Ανάμεσα στους δύο διαιρετικούς αλγορίθμους, K-Medoids και K- Modes, επιλέχτηκε ο K- Modes γιατί έχει καλύτερα scores στη μέθοδο του Silhouette score. Από τη γραφική παράσταση του K- Modes (Εικόνα 23) διαπιστώνεται ότι για 2 ή 7 συστάδες προκύπτει το καλύτερο αποτέλεσμα. Επιλέχτηκε ο αριθμός 7 γιατί στην περίπτωση για 2 συστάδες δεν παρουσιάστηκαν χαρακτηριστικές ομάδες αγοραστικών αποστόλων και επιπλέον ο αριθμός 7 εμφανίστηκε και στη γραφική παράσταση της μεθόδου «Elbow method».

4.4.3 Κατάτμηση Επισκέψεων (Visits Segmentation) Evaluation Step

Ο K-Mode όπως έχει αναφερθεί, είναι κατάλληλος αλγόριθμος για την ομαδοποίηση δυαδικών μεταβλητών. Με μέτρο, την απόσταση Jaccard, οι επισκέψεις που πραγματοποιήθηκαν, ομαδοποιήθηκαν βάσει των κοινών περιοχών τους. Σ' αυτό το βήμα της ανάλυσης, βάσει της Crisp-DM μεθοδολογίας γίνεται η αξιολόγηση των αποτελεσμάτων και αξιολογείται κατά πόσο τα αποτελέσματα του μοντέλου που σχεδιάστηκε, έχουν αξία για την επιχείρηση.

Μετά την κατασκευή του μοντέλου, βάσει των τεχνικών αξιολόγησης της μοντελοποίησης που παρουσιάστηκαν στο προηγούμενο βήμα, τα αποτελέσματα θα πρέπει να αξιολογηθούν και από τη σκοπιά των επιχειρησιακών στόχων. Σκοπός είναι να καθοριστεί η χρησιμότητα της εξαγχθείσας γνώσης στη λύση του προβλήματος που είχε τεθεί αρχικά.

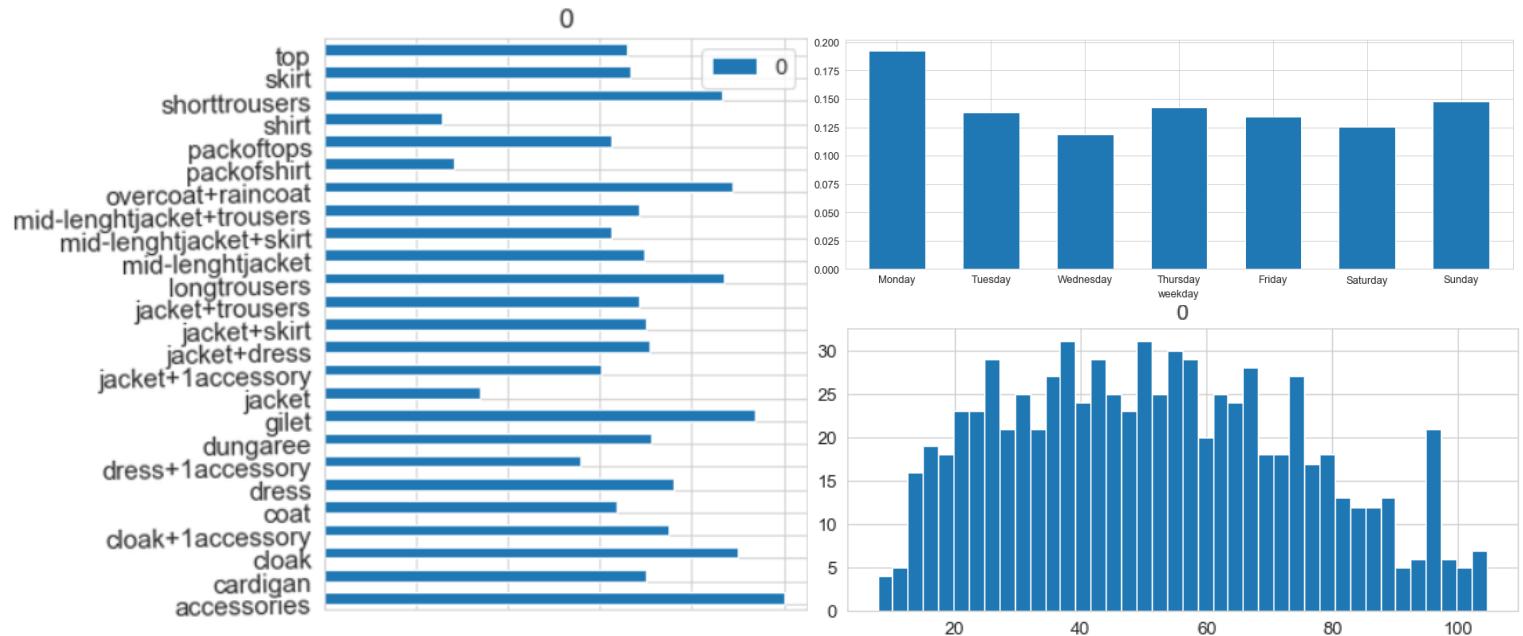
Τα αποτελέσματα της κατάτμησης με τον K-Modes, για 7 συστάδες παρουσιάζονται παρακάτω:



Εικόνα 24: Αριθμός επισκέψεων σε κάθε συστάδα

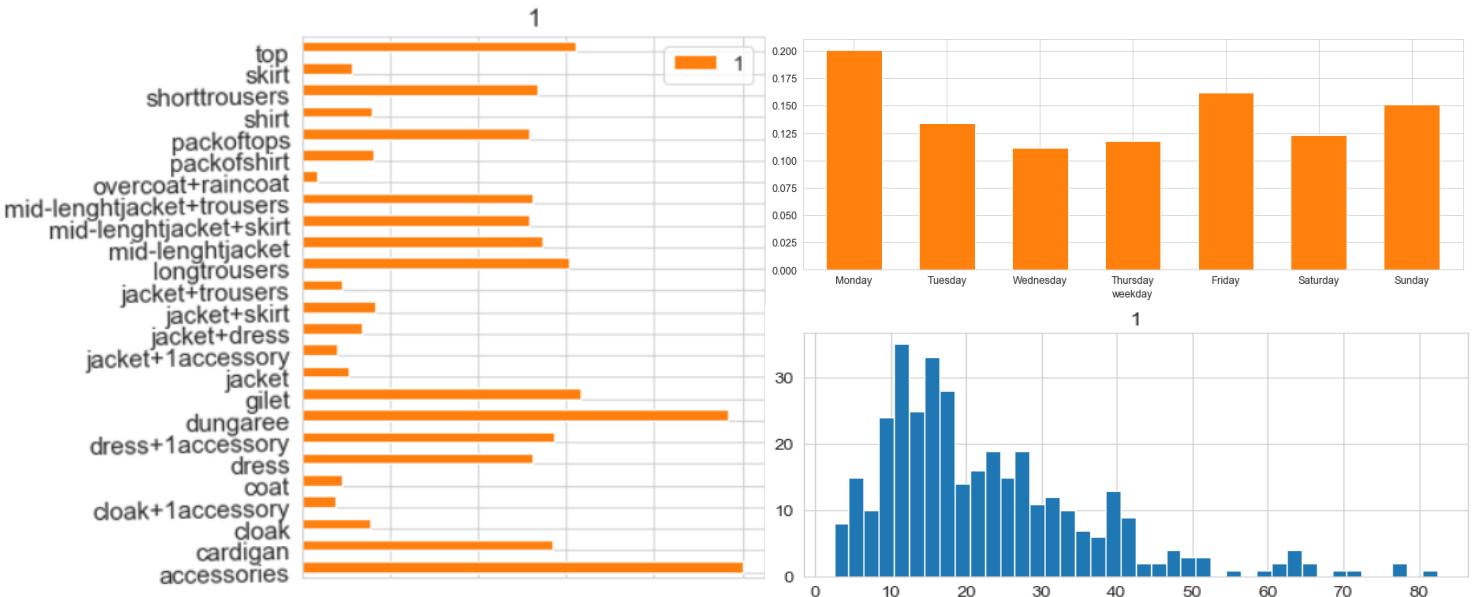
Mean values				
cluster	category	session_duration	dist_category	hourin
0	48.961190	51.849547	18.103493	10.538163
1	22.192737	22.848650	10.508380	10.840782
2	46.176252	48.852412	18.172542	10.749536
3	15.742537	15.936505	8.373134	11.832090
4	13.123967	13.399725	7.074380	11.628099
5	20.312950	20.584293	10.708633	10.392086
6	24.336806	25.252373	11.770833	10.795139

Πίνακας 7: Συγκεντρωτικά στατιστικά των ομάδων επισκέψεων



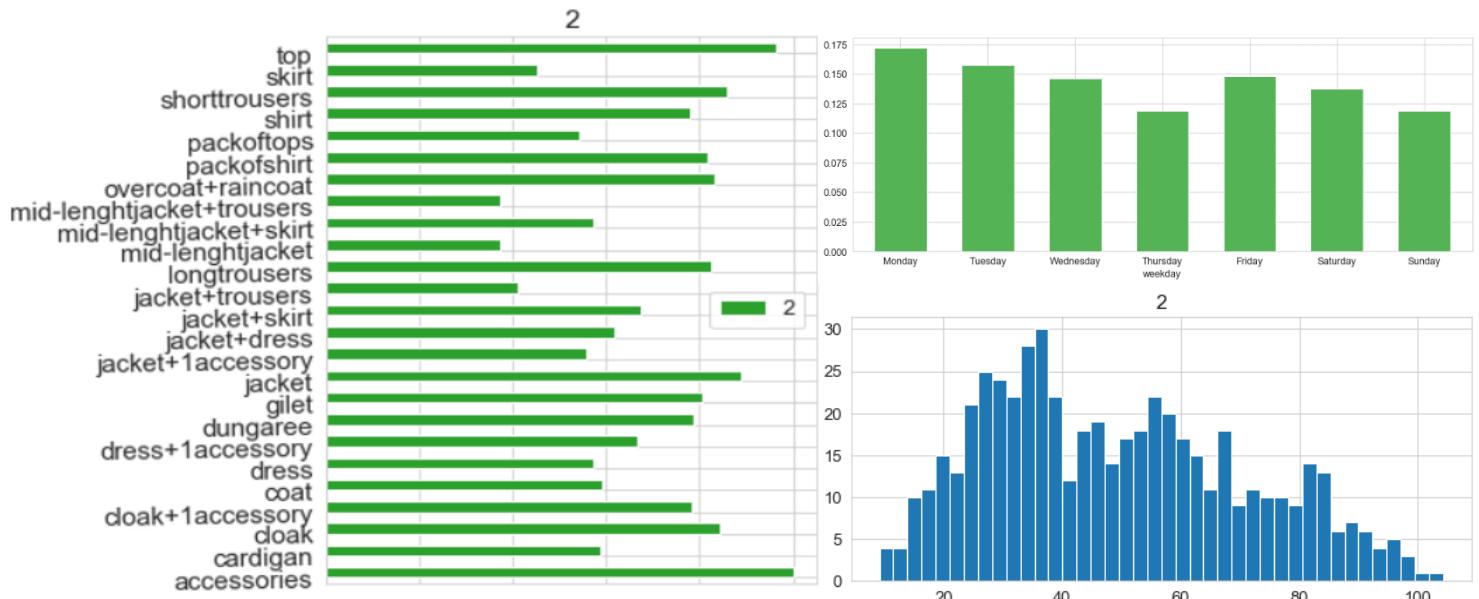
Εικόνα 25: Περιγραφικά στατιστικά για την πρώτη ομάδα επισκέψεων

Η συστάδα 0 αποτελείται από 773 επισκέψεις (29% των συνολικών επισκέψεων). Από τις παραπάνω γραφικές παραστάσεις φαίνεται ότι η συστάδα 0 πήγε σε όλες τις κατηγορίες προϊόντων εκτός από τρεις. Αυτή η συστάδα, κατά μέσο όρο, επισκέφτηκε 18 περιοχές. Ο χρόνος παραμονής των πελατών στο κατάστημα μοιάζει με κανονική κατανομή, που σημαίνει ότι υπήρχαν και πελάτες που ψώνισαν πολύ γρήγορα και πελάτες που περιτριγύρισαν αρκετή ώρα μέσα σε αυτό, και με την πλειοψηφία των πελατών να βρίσκεται γύρω από το μέσο όρο της διάρκειας επίσκεψης, που είναι περίπου 50 λεπτά. Η ημέρα με τις περισσότερες επισκέψεις ήταν η Δευτέρα με 20% και η ημέρα με τις λιγότερες, ήταν η Τετάρτη με 12%.



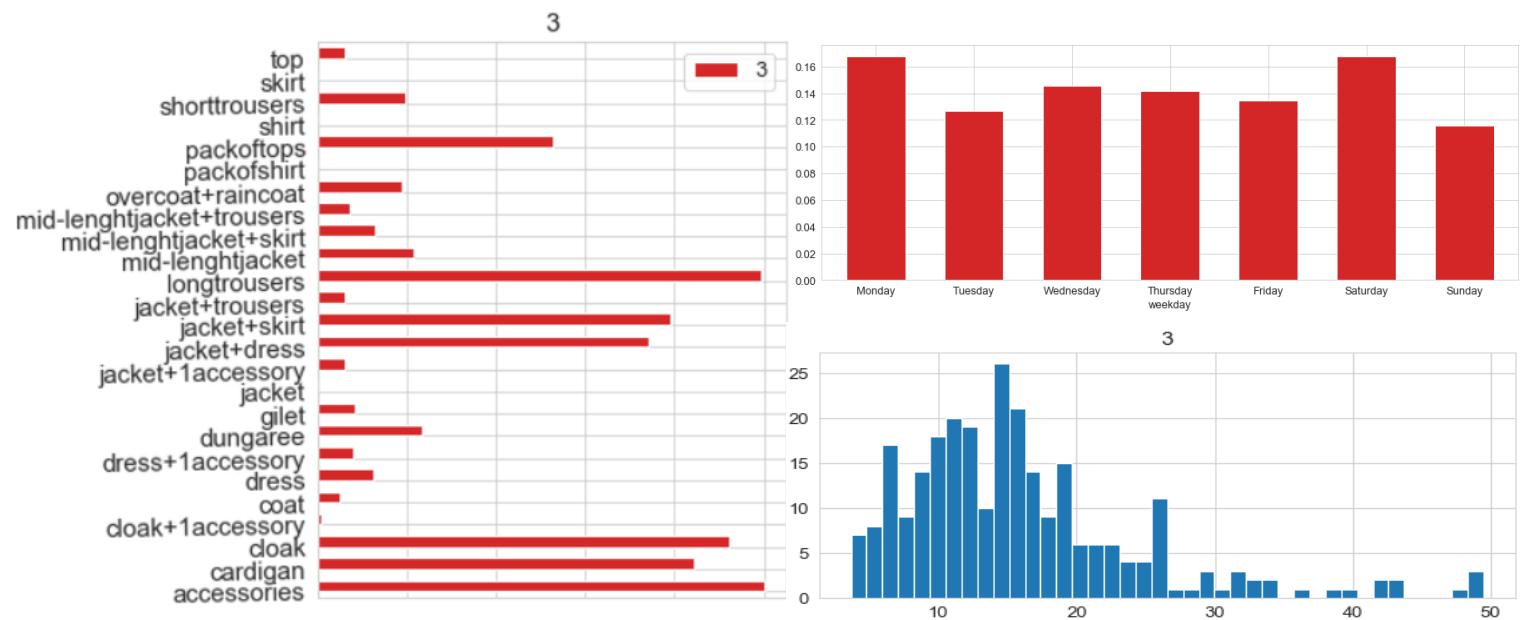
Εικόνα 26: Περιγραφικά στατιστικά για τη δεύτερη ομάδα επισκέψεων

Η συστάδα 1 αποτελείται από 358 επισκέψεις (13% των συνολικών επισκέψεων). Από τις παραπάνω γραφικές παραστάσεις φαίνεται ότι συστάδα 1, κατά πλειοψηφία, πήγε στην κατηγορία προϊόντων dungaree, η οποία περιοχή βρίσκεται κοντά στη μια από τις δύο εισόδους του καταστήματος. Αυτή η συστάδα, κατά μέσο όρο, επισκέφτηκε 10 περιοχές και έδειξε ενδιαφέρον για τα προϊόντα που βρίσκονται γύρω από την περιοχή με τα mid-length jacket. Η γραφική παράσταση με το χρόνο, δείχνει ότι η πλειοψηφία των πελατών αυτής της συστάδας παρέμεινε στο κατάστημα για μικρό χρονικό διάστημα, κατά μέσο όρο 22 λεπτά. Η ημέρα με τις περισσότερες επισκέψεις ήταν η Δευτέρα με 20% και η ημέρα με τις λιγότερες, ήταν η Τετάρτη με 11%.



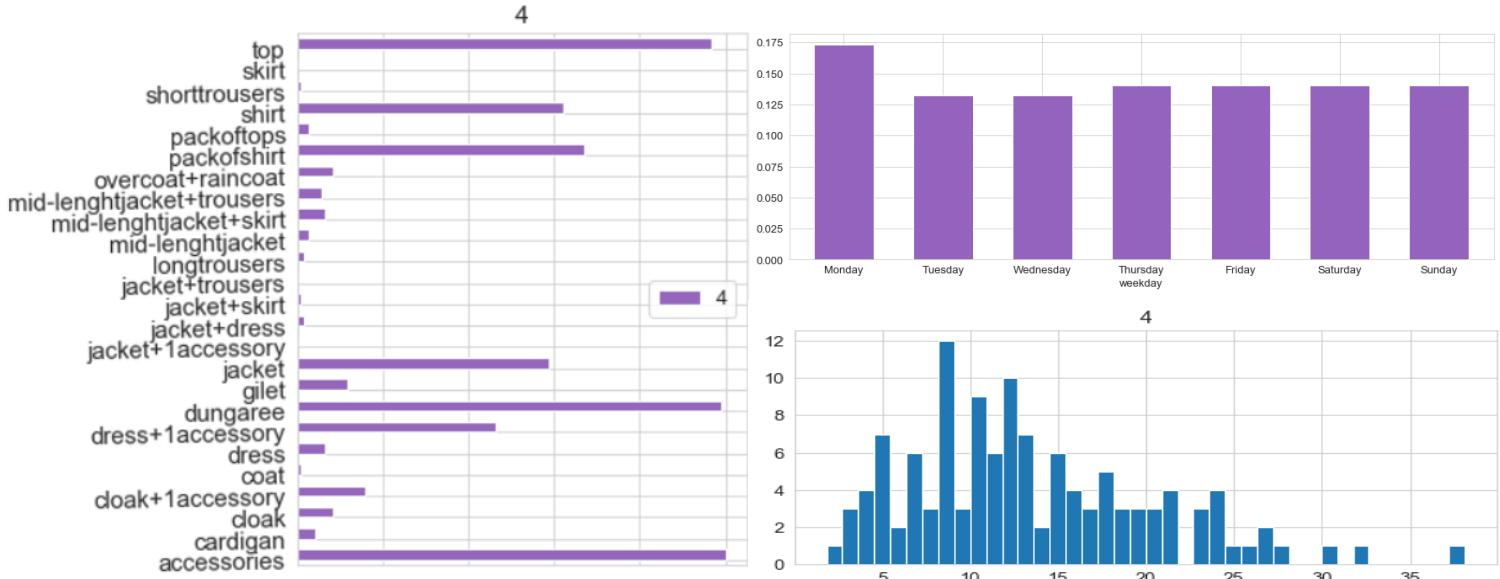
Εικόνα 27: Περιγραφικά στατιστικά για την τρίτη ομάδα επισκέψεων

Η συστάδα 2 αποτελείται από 539 επισκέψεις (20% των συνολικών επισκέψεων). Από τις παραπάνω γραφικές παραστάσεις φαίνεται ότι η συστάδα 2 πήγε σε όλες τις κατηγορίες προϊόντων εκτός από τέσσερις. Αυτή η συστάδα, κατά μέσο όρο, επισκέφτηκε 18 περιοχές. Ο μέσος όρος της διάρκειας επίσκεψης, είναι περίπου 48 λεπτά. Η ημέρα με τις περισσότερες επισκέψεις ήταν η Δευτέρα με 17% και η ημέρα με τις λιγότερες, ήταν η Πέμπτη με 12%.



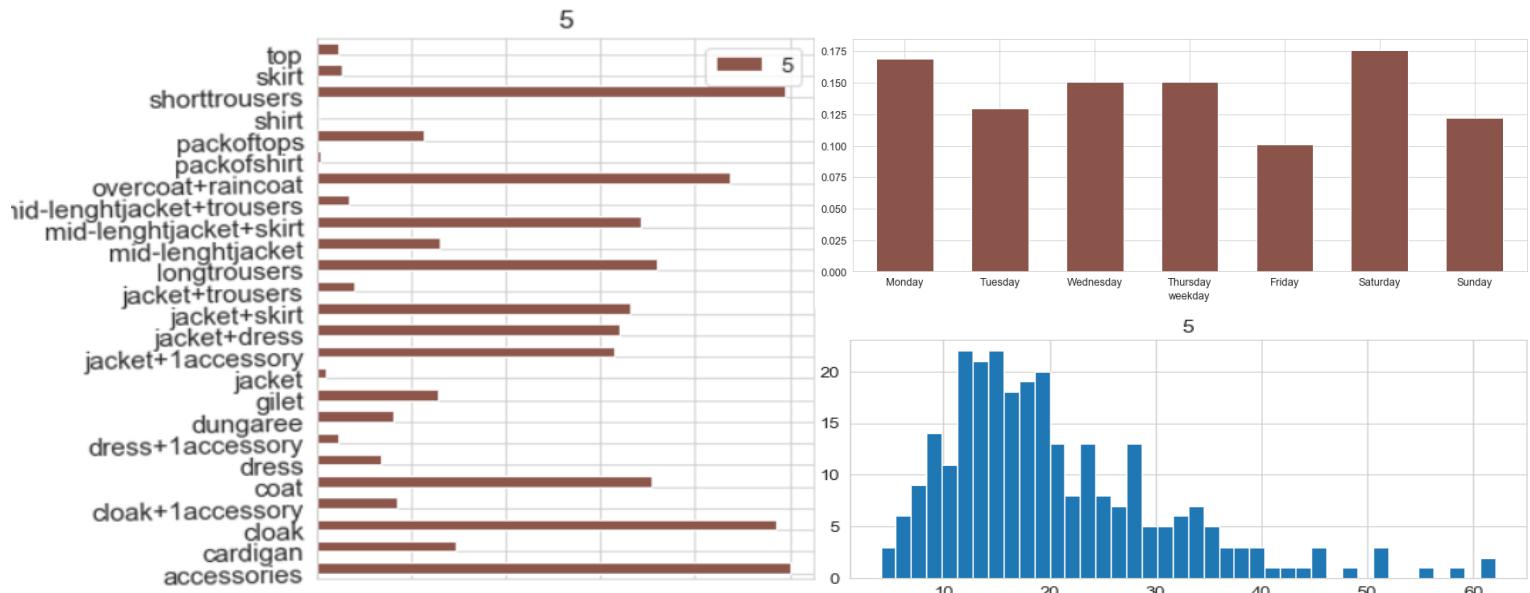
Εικόνα 28: Περιγραφικά στατιστικά για την τέταρτη ομάδα επισκέψεων

Η συστάδα 3 αποτελείται από 268 επισκέψεις (10% των συνολικών επισκέψεων). Από τις παραπάνω γραφικές παραστάσεις φαίνεται ότι συστάδα 3, κατά πλειοψηφία, πήγε στην κατηγορία προϊόντων cloak, η οποία περιοχή βρίσκεται κοντά στη μια από τις δύο εισόδους του καταστήματος, που σημαίνει ότι οι πελάτες μπήκαν από αυτήν την ίδια είσοδο. Επίσης παρατηρείται ότι κατά μέσο όρο πήγε σε 8 κατηγορίες προϊόντων ,από τις οποίες οι πιο χαρακτηριστικές είναι παρακάτω 7: cloak, cardigan, long trousers, packoftops, jacket+dress, jacket+skirt, accessories. Η γραφική παράσταση με το χρόνο, δείχνει ότι η πλειοψηφία των πελατών αυτής της συστάδας παρέμεινε στο κατάστημα για μικρό χρονικό διάστημα, κατά μέσο όρο 16 λεπτά. Οι ημέρες με τις περισσότερες επισκέψεις ήταν η Δευτέρα και το Σάββατο με 16% και η ημέρα με τις λιγότερες, ήταν η Κυριακή με 11%.



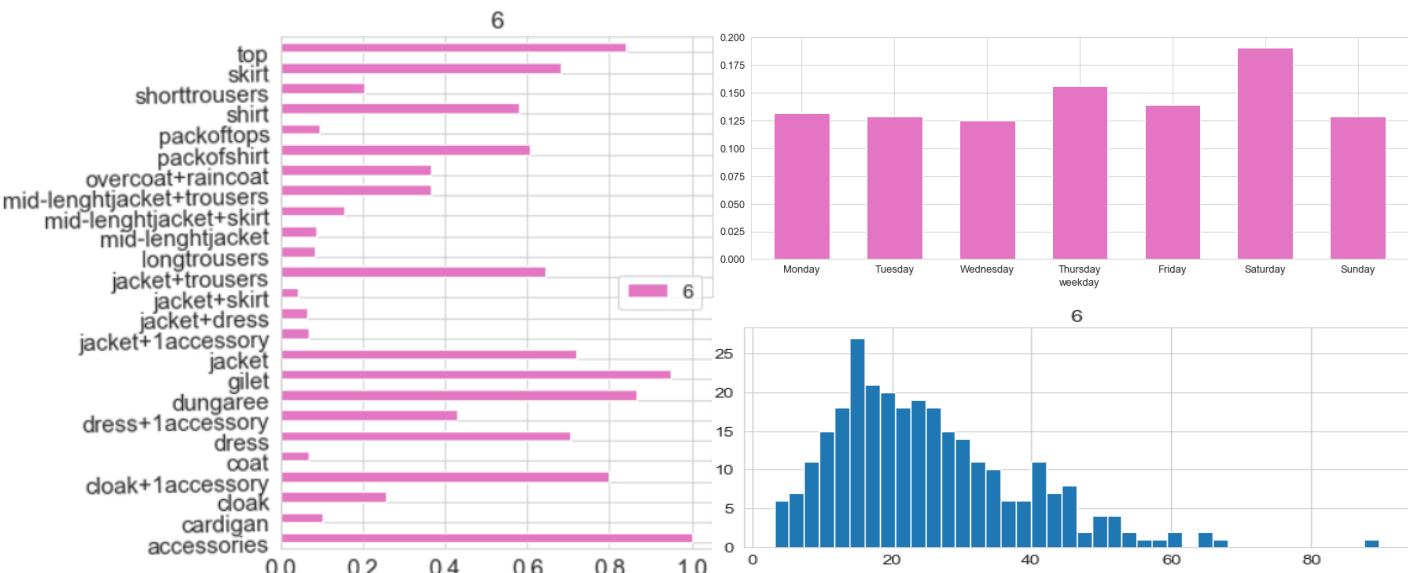
Εικόνα 29: Περιγραφικά στατιστικά για την πέμπτη ομάδα επισκέψεων

Η συστάδα 4 αποτελείται από 121 επισκέψεις (4% των συνολικών επισκέψεων). Από τις παραπάνω γραφικές παραστάσεις φαίνεται ότι συστάδα 4, κατά πλειοψηφία, επισκέφτηκε την κατηγορία προϊόντων dungaree, η οποία περιοχή βρίσκεται κοντά στη μια από τις δύο εισόδους του καταστήματος, που σημαίνει ότι οι πελάτες αυτής της συστάδας, μπήκαν από αυτήν την ίδια είσοδο. Επίσης παρατηρείται ότι οι πελάτες, κατά μέσο όρο πήγαν σε 7 κατηγορίες προϊόντων, από τις οποίες οι πιο χαρακτηριστικές είναι παρακάτω 6: dungaree, packofshirt, jacket, shirt, top, accessories. Η γραφική παράσταση με το χρόνο παραμονής, δείχνει ότι η πλειοψηφία των πελατών αυτής της συστάδας παρέμεινε στο κατάστημα, το λιγότερο χρόνο από όλες τις συστάδες, κατά μέσο όρο 13 λεπτά. Η ημέρα με τις περισσότερες επισκέψεις ήταν η Δευτέρα με 17% και οι υπόλοιπες ημέρες παρουσίασαν την ίδια επισκεψιμότητα γύρω στο 13%.



Εικόνα 30: Περιγραφικά στατιστικά για την έκτη ομάδα επισκέψεων

Η συστάδα 5 αποτελείται από 278 επισκέψεις (10% των συνολικών επισκέψεων). Από τις παραπάνω γραφικές παραστάσεις φαίνεται ότι η συστάδα 5, κατά πλειοψηφία, επισκέφτηκε την κατηγορία προϊόντων cloak, η οποία περιοχή βρίσκεται κοντά στη μια από τις δύο εισόδους του καταστήματος, που σημαίνει ότι οι πελάτες αυτής της συστάδας, μπήκαν από αυτήν την ίδια είσοδο. Επίσης παρατηρείται ότι οι πελάτες, κατά μέσο όρο πήγαν σε 11 κατηγορίες προϊόντων και έδειξαν ενδιαφέρον για τα προϊόντα που βρίσκονται γύρω από την περιοχή με τα jacket και τους ενδυματολογικούς συνδυασμούς του. Η γραφική παράσταση με το χρόνο, δείχνει ότι η πλειοψηφία των πελατών αυτής της συστάδας παρέμεινε στο κατάστημα για ένα χρονικό διάστημα, κατά μέσο όρο 20 λεπτά. Η ημέρα με τις περισσότερες επισκέψεις ήταν το Σάββατο με 17,5% και η ημέρα με τις λιγότερες, ήταν η Παρασκευή με 10%.



Εικόνα 31: Περιγραφικά στατιστικά για την έθδομη ομάδα επισκέψεων

Η συστάδα 6 αποτελείται από 288 επισκέψεις (10% των συνολικών επισκέψεων). Από τις παραπάνω γραφικές παραστάσεις φαίνεται ότι η συστάδα 6 επισκέφτηκε την κατηγορία προϊόντων dungaree, η οποία περιοχή βρίσκεται κοντά στη μια από τις δύο εισόδους του καταστήματος, που υποδεικνύει ότι οι πελάτες αυτής της συστάδας, μπήκαν από αυτήν την ίδια είσοδο. Επίσης παρατηρείται ότι οι πελάτες, κατά μέσο όρο πήγαν σε 11 κατηγορίες προϊόντων και έδειξαν ενδιαφέρον για τα προϊόντα όπως gilet και top. Η γραφική παράσταση με το χρόνο, δείχνει ότι η πλειοψηφία των πελατών αυτής της συστάδας παρέμεινε στο κατάστημα για ένα χρονικό διάστημα, κατά μέσο όρο 25 λεπτά. Οι ημέρες με τις περισσότερες επισκέψεις ήταν το Σάββατο με 19% και η Πέμπτη με 16% και η ημέρα με τις λιγότερες, ήταν η Τετάρτη με 12,5%.

Συμπεράσματα της Ανάλυσης

Συγκεντρωτικά, μελετώντας τα αποτελέσματα της ανάλυσης των δεδομένων, όπως αυτά φαίνονται στον παραπάνω πίνακα και στις παραπάνω εικόνες, μπορούν να εξαχθούν τα παρακάτω συμπεράσματα:

Οι κατηγορίες επισκέψεων είναι άρρηκτα συνδεδεμένες με τη χρονική τους διάρκεια, όπως έχει αναφερθεί και προηγουμένως. Έτσι οι ομαδοποιήσεις που προκύπτουν από τις κατηγορίες, μπορούν να χαρακτηριστούν έμμεσα και από τη χρονική διάρκεια, χωρίς αυτή η μεταβλητή να έχει εισαχθεί στο μοντέλο και αυτό μπορεί να δώσει επιπλέον συμπεράσματα για τις συστάδες. Οι ομάδες 3, 4 αποτελούνται από μικρής διάρκειας επισκέψεις. Οι 1, 5, 6 ομάδες από μέσης διάρκειας επισκέψεις και οι 0, 2 ομάδες από μακράς διάρκειας επισκέψεις (κατά μέσο όρο 50 λεπτών).

Να σημειωθεί ότι έγινε προσπάθεια να μελετηθούν οι επισκέψεις βάσει της χρονικής τους διάρκειας, με το να χωριστούν οι επισκέψεις σε 3 υποκατηγορίες βάσει της διάρκειάς τους. Για αυτό το σκοπό χρησιμοποιήθηκε ο K-Means με k=3 στο προπαρασκευαστικό στάδιο της επεξεργασίας των δεδομένων με μόνη μετρική σύγκρισης τη συνολική διάρκεια της κάθε επίσκεψης και οι επισκέψεις διαχωρίστηκαν σε 3 υποκατηγορίες. Σε κάθε ένα από αυτά τα 3 σετ δεδομένων εφαρμόστηκε ο αλγόριθμος των K-modes (μέθοδος 2 step clustering). Η ανάλυση αυτή κατέληξε σε παρόμοια αποτελέσματα με την αντιστοίχιση, όπως παρουσιάστηκε στην παραπάνω παράγραφο.

Κάποιες συστάδες, μπορούν να ερμηνευτούν από ομάδες πελατών που μπήκαν στο κατάστημα με συγκεκριμένους στόχους αγοράς (shopping missions). Οι πελάτες αυτών των συστάδων ήζεραν τι ήθελαν να αγοράσουν και αυτό φαίνεται από τα ποσοστά επίσκεψης στις χαρακτηριστικές περιοχές της αντίστοιχης συστάδας.

Υπάρχουν συστάδες που δημιουργήθηκαν από τον αλγόριθμο ακολουθώντας τους χωρικούς περιορισμούς του καταστήματος. Σε συνδυασμό με το χαρακτηριστικό ότι οι επισκέψεις είχαν μικρά μονοπάτια, σημαίνει ότι οι πελάτες που μπήκαν από την ίδια είσοδο έκαναν παρόμοιες κλειστές διαδρομές. Μια παραπάνω υπόθεση είναι ότι η είσοδος του καταστήματος που μπήκε ο πελάτης, καθόρισε τη διαδρομή του.

Αυτή η μέθοδος ανάλυσης δεδομένων μπορεί να βοηθήσει τους λιανέμπορους να οργανώσουν και να σχεδιάσουν κατάλληλα τις προωθητικές και διαφημιστικές τους ενέργειες (προώθηση, προσφορές) εντός και εκτός του καταστήματος, για να έχουν τα μέγιστα αποτελέσματα, δηλαδή



την αύξηση των αυθόρμητων αγορών και των παράλληλων πωλήσεων σε συγκεκριμένα προϊόντα τα οποία θα αποφασίσουν πότε (ποιες μέρες) και πού θα τοποθετηθούν (σημεία ενδιαφέροντος).

Για το λόγο αυτό, οι λιανέμποροι, καλό είναι να συμπεριλάβουν στη μελέτη των δράσεων τους την επισκεψιμότητα ανά ημέρα και ανά περιοχή του καταστήματος.

	0	1	2	3	4	5	6
accessories	100	100	100	100	100	100	100
cardigan	69	56	58	84	4	29	10
cloak	89	15	84	92	8	97	25
cloak+1accessory	75	7	78	0	15	16	79
coat	63	8	59	4	0	70	6
dress	76	52	57	12	6	13	70
dress+1accessory	55	57	66	7	46	4	43
dungaree	71	96	78	23	99	16	86
gilet	93	63	80	8	11	25	94
jacket	33	10	88	0	58	1	71
jacket+1accessory	60	8	55	5	0	62	6
jacket+dress	70	13	61	74	1	64	6
jacket+skirt	70	16	67	79	0	66	4
jacket+trousers	68	9	41	5	0	7	64
longtrousers	86	60	82	99	1	71	8
mid-lengthjacket	69	54	37	21	2	25	8
mid-lengthjacket+skirt	62	51	57	12	6	68	15
mid-lengthjacket+trousers	68	52	37	7	5	6	36
overcoat+raincoat	88	3	83	19	8	87	36
packofshirt	28	16	81	0	66	0	60
packoftops	62	51	54	52	2	22	9
shirt	25	15	78	0	61	0	57
shorttrousers	86	53	85	19	0	98	20
skirt	66	11	45	0	0	5	68
top	65	62	96	5	96	4	84

Εικόνα 32: Συγκεντρωτικά τα ποσοστά επίσκεψης ανά κατηγορία και συστάδα

5

Managerial implications

Η ανάλυση δεδομένων πελατών (Web Analytics), είναι μια καθιερωμένη πρακτική στο διαδικτυακό λιανεμπόριο. Τα δεδομένα πελατών που συλλέγονται μέσω του προγράμματος περιήγησης, συγκρίνονται με τα δεδομένα συναλλαγών που είναι αποθηκευμένα στο σύστημα αγορών, και έτσι οι συναλλαγές συγκρίνονται με τις πιθανές συναλλαγές (επισκέπτες του ιστότοπου) στο λεγόμενο ποσοστό μετατροπής (conversion rate). Το Web-Analytics είναι μια μέθοδος χρήσης των δεδομένων του προγράμματος περιήγησης, που λαμβάνονται για τη βελτίωση της δομής, της ρύθμισης της σελίδας προορισμού και της τιμολόγησης, καθώς και των προσφορών ενός διαδικτυακού καταστήματος, με τρόπο που να αυξάνει τις πωλήσεις. Αυτή η διαδικτυακή ανάλυση αποτελεί ουσιαστικό μέρος σχεδόν όλων των διαδικτυακών πλατφορμών. Σ' αντίθεση με αυτό, τα δεδομένα πελατών για επισκέψεις στο κατάστημα, αποτελούν μια σχεδόν ανεξερεύνητη περιοχή για τις επιχειρήσεις λιανικού εμπορίου. Παρόλο που έχουν αναπτυχθεί και αναλυθεί στρατηγικές για την ανάλυση δεδομένων επισκεπτών για δεκαετίες (π.χ. μέσω μη αυτόματης καταμέτρησης, χρήσης δοκιμαστικών πελατών, χρήσης παρακολούθησης ματιών ή χρήσης συστημάτων μοντελοποίησης όπως blueprinting), η (εν μέρει) αυτοματοποιημένη αξιολόγηση και ανάλυση δεδομένων πελατών (In Store Customer Analytics) δεν έχει γίνει ακόμη πάγια πρακτική. Αντ' αυτού, οι έμποροι λιανικής και οι ερευνητές, συνεχίζουν να επικεντρώνονται στην ανάλυση δεδομένων συναλλαγών και έτσι παραμελούν την ευκαιρία να επεκτείνουν τον ορίζοντα ανάλυσής τους, ώστε να συμπεριλάβουν δεδομένα πελατών εντός καταστήματος που δεν καλύπτονται από τις συναλλαγές.

Συνεπώς, στην πρώτη ενότητα αυτού του κεφαλαίου θα γίνει αναφορά στους δείκτες απόδοσης KPIs, οι οποίοι εφαρμόζονται με επιτυχία στα ψηφιακά καταστήματα και χρησιμοποιούνται συχνά από τους λιανέμπορους για τη λήψη αποφάσεων. Στη συνέχεια αυτής της ενότητας θα γίνει παρουσίαση της εφαρμογής των KPIs στην ανάλυση της απόδοσης του φυσικού καταστήματος, της παρούσας εργασίας.

Στη δεύτερη ενότητα θα παρουσιαστούν οι δυνατότητες εφαρμογής της τεχνολογίας συλλογής δεδομένων θέσης (BLEs) η οποία επιλέχθηκε για τη συλλογή των δεδομένων της εργασίας και πως μπορούν να ωφεληθούν οι εταιρείες λιανεμπορίου από τη χρήση της (π.χ. ορθή λήψη αποφάσεων).

5.1 Key Performance Indicators (KPIs)

Για να μπορεί μια επιχείρηση να αξιολογήσει την απόδοση της ή να εξακριβώσει αν επιτευχθήκαν οι στρατηγικοί της στόχοι, θα πρέπει να συλλέγει δεδομένα, να τα επεξεργάζεται και να τα αναλύει, έτσι ώστε να αποφασίσει τα επόμενα βήματα δράσης της. Με την αξιολόγηση διαπιστώνονται οι παράγοντες που μεγιστοποιούν την απόδοση της επένδυσης (Return on Investment - ROI), τι φέρνει θετικά αποτελέσματα στην κερδοφορία και την εξέλιξη της επιχείρησης και τι δεν αποδίδει ικανοποιητικά, έτσι ώστε να ανακατανείμει του πόρους της σε άλλες ενέργειες ή καμπάνιες που μπορεί να αποδώσουν σε μεγαλύτερο βαθμό (Saura et al., 2017). Για αυτό τον λόγο, υπάρχουν οι Βασικοί Δείκτες Απόδοσης (Key Performance Indicator – KPIs), που είναι μετρήσιες που βοηθούν στην εξαγωγή συμπερασμάτων για το κατά ποσό έχουν επιτευχθεί οι επιχειρησιακοί στόχοι που έχει θέσει η επιχείρηση (Sharma, 2015). Άρα, αφού διευκρινιστούν και αποφασιστούν ποιοι στόχοι είναι προς επίτευξη και σημαντικοί για την κερδοφορία της επιχείρησης, έπειτα, με την χρήση των KPIs, λαμβάνονται τα μετρήσιμα αποτελέσματα τα οποία καταδεικνύουν σε ποιον βαθμό επιτεύχθηκαν οι στόχοι που είχαν τεθεί από την επιχείρηση. Δεδομένου ότι τα KPIs είναι μετρήσιες, συνήθως είναι αριθμοί ή ποσοστά. Σύμφωνα με τους (Kumar et al., 2012) βασικά χαρακτηριστικά των KPIs είναι τα κάτωθι:

- Να είναι μετρήσιμοι με αριθμούς και να είναι σχετικοί με το αντικείμενο που εξετάζεται, ώστε να έχουν αξία για την επιχείρηση
- Είναι σημαντικό να μπορεί να εξάγεται γρήγορα η πληροφορία που δίνει ο συγκεκριμένος δείκτης, έτσι ώστε να μπορεί να υπολογίσει εύκολα την απόκλιση στα δεδομένα, μόλις τον δει ο χρήστης
- Να είναι αντιληπτές από άτομα με διαφορετικό υπόβαθρο, έτσι ώστε σε μια εταιρεία, άτομα από διαφορετικά τμήματα να μπορούν να λαμβάνουν αποφάσεις εύκολα, αναλύοντας τις μετρικές
- Να είναι χρονικά επίκαιροι. Δηλαδή, πρέπει να παρέχονται άμεσα οι μετρικές, ώστε οι υπεύθυνοι λήψης αποφάσεων να μπορούν να λαμβάνουν έγκαιρα αποφάσεις. Καθίστανται αναποτελεσματικοί, αν απαιτούν π.χ. ένα μήνα για την εξαγωγή χρήσιμων πληροφοριών από αυτούς, όταν η βιομηχανία αλλάζει σε μια εβδομάδα. Σε αυτή την περίπτωση, η συγκεκριμένη μετρική όσο εξαιρετική και αν είναι, καθίσταται μη ιδανική για εξαγωγή συμπερασμάτων.



5.1.1 Ορισμός Και Αξιοποίηση Των Key Performance Indicators (Kpis)

Αριθμός των Επισκεπτών ανά Κατηγορία

Η σημαντικότερη μετρική που θα αναφερθεί, είναι ο μέσος αριθμός των επισκεπτών ανά κατηγορία του καταστήματος. Αυτή η μετρική μπορεί να δώσει στους μάνατζερς πληροφορία για την κινητικότητα των καταναλωτών στο κατάστημα. Και από αυτή την μετρική μπορεί να μετρηθεί η απόδοση των ενεργειών του τμήματος μάρκετινγκ της εταιρείας μέσα στο κατάστημα. Αν για παράδειγμα θέλει ο λιανέμπορος να βελτιώσει τις στρατηγικές των προωθητικών ενεργειών εντός του καταστήματος και την επικοινωνία με τον πελάτη, χρειάζεται να ξέρει πως κινείται ο πελάτης και σε ποιες περιοχές του καταστήματος πάει. Έτσι μπορεί να αξιολογήσει αν οι πελάτες ανταποκρίνονται θετικά στις προωθητικές ενέργειες. Μέσω ενός a/b test, ο λιανέμπορος μπορεί να δει αν αυξήθηκαν οι επισκέψεις σε μια περιοχή, η οποία είναι υπό διαφήμιση για ένα χρονικό διάστημα, πριν και μετά τη διαφήμιση και με έλεγχο υποθέσεων να εξετάσει αν είναι στατιστικά σημαντική η αύξηση ή η μείωση της επισκεψιμότητας στη συγκεκριμένη περιοχή. Αυτόν τον έλεγχο ο λιανέμπορος μπορεί να το κάνει έμμεσα, με το να παρατηρήσει τη μεταβολή στις πωλήσεις, άλλα ο δείκτης πωλήσεων δεν δείχνει όλη την εικόνα, δηλαδή τους πελάτες που όντως επισκέφτηκαν την περιοχή με πρόθεση να αγοράσουν, άλλα εν τέλει δεν προέβησαν σε αγορά.

Αυτή η μετρική είναι επίσης σημαντική για τη διαμόρφωση του καταστήματος και για την αύξηση παρορμητικών αγορών (Herb Sorensen et al., 2017). Άλλες έρευνες που μελέτησαν την κίνηση των καταναλωτών μέσα σε καταστήματα και έχουν σχέση με αυτή την μετρική, έδειξαν ότι το ποσοστό των πελατών που επισκέπτονται όλο το κατάστημα είναι πολύ μικρό (Herb Sorensen, 2009). Βασικός στόχος του μάνατζερ είναι να ταυτοποιήσει ποιες περιοχές είναι οι λιγότερο επισκεπτόμενες (και τι μπορεί να κάνει για αυτό) και ποιες είναι οι «σημαντικές», δηλαδή οι πιο συχνά επισκεπτόμενες περιοχές.

Browsing Κατηγορία

Πόσες φορές επισκέπτεται ο καταναλωτής, σε μια επίσκεψη (session), κάθε κατηγορία. Και μετά υπολογίζεται ο μέσος όρος για κάθε κατηγορία. Ο δείκτης αυτός δείχνει κατά πόσες φορές ο καταναλωτής επέστρεψε σε μια κατηγορία προϊόντων. Είναι μια ένδειξη για το αν ο καταναλωτής έχει ένα συγκεκριμένο αγοραστικό σκοπό και επιστρέφει για να βρει κάτι συγκεκριμένο ή απλά κάνει έρευνα αγοράς. Σε συνεργασία με τα δεδομένα POS, δίνεται μια εικόνα αν τα προϊόντα είναι σωστά τοποθετημένα, δηλαδή μπορεί να διαπιστωθεί αν οι καταναλωτές μπορούν να βρουν εύκολα αυτό που ψάχνουν.



Μέσος Χρόνος Επίσκεψης ανά Κατηγορία

Άλλη σημαντική μετρική, είναι ο μέσος χρόνος επίσκεψης σε κάθε κατηγορία. Ο χρόνος που περνάει ο καταναλωτής μέσα στο κατάστημα, επηρεάζει το πως κινείται και σε τι αγορές προβαίνει. Όταν οι λιανέμποροι αναλύουν τη συμπεριφορά των καταναλωτών για να εφαρμόσουν πιο αποδοτικές στρατηγικές μάρκετινγκ, παρατηρούν και το χρόνο παραμονής στο κατάστημα (Herb Sorensen et al., 2017). Με αυτήν την μετρική έχουμε αυτή την πληροφορία σε επίπεδο κατηγορίας. Και όπως αναφέρεται και στη βιβλιογραφία, όσο περισσότερο χρόνο δαπανά ένας πελάτης σε ένα κατάστημα τόσο αυξάνεται η πιθανότητα να προβεί σε αγορά (Ferracuti et al., 2019).

Έχει παρατηρηθεί ότι ένα μεγάλο κομμάτι του χρόνου παραμονής των πελατών μέσα στο κατάστημα, δεν σχετίζεται με τις αγορές τους. (Hui et al., 2009) Για αυτό το λόγο, από τη μεριά του λιανέμπορου είναι σημαντικό να γνωρίζει σε ποιες περιοχές οι πελάτες παραμένουν και από ποιες περιοχές απλά διέρχονται. Για να παρέχουν αυτή την πληροφορία στο λιανέμπορο, οι Ferracuti et al αναφέρουν το δείκτη απόδοσης GRP (Gross Rating Points). Αυτός ο δείκτης προκύπτει από τον πολλαπλασιασμό των δυο παραπάνω μετρικών, του χρόνου παραμονής στην περιοχή και του πλήθους των πελατών.

Ποσοστό Επίσκεψης ανά Κατηγορία

Είναι το ποσοστό του χρόνου που πέρασε ο καταναλωτής συνολικά ανά συγκεκριμένη κατηγορία, προς τον συνολικό χρόνο της επίσκεψης (session) του καταναλωτή καθώς και το μέσο ποσοστό του χρόνου που δαπάνησαν οι πελάτες ανά επίσκεψη στην αντίστοιχη κατηγορία. Δηλαδή ο δείκτης αυτός δείχνει το ποσοστό του χρόνου της επίσκεψης, που δαπανούν οι καταναλωτές ανά κατηγορία προϊόντων και εκφράζεται σε ποσοστό %. Είναι ενδιαφέρον για τη διοίκηση της επιχείρησης, να συγκριθεί ο χρόνος που περνάνε οι πελάτες στο κατάστημα ανά κατηγορία προϊόντων. Και πιο σημαντικό ακόμη είναι να παρατηρηθεί η χρονική διακύμανση του δείκτη στο χρόνο (π.χ. μήνα ή τρίμηνο κ.λ.π.)

Σειρά Προτεραιότητας

Ο δείκτης αυτός δείχνει με ποια σειρά επισκέπτονται οι πελάτες τις διάφορες κατηγορίες προϊόντων. Δηλαδή, τι προτεραιότητα έχει μια κατηγορία στην επίσκεψη ενός καταναλωτή, και κατά πόσο είναι σημαντική γι' αυτόν (shopping mission). Η σημαντικότητα μιας κατηγορίας προϊόντων, εξαρτάται από τη θέση της κατηγορίας στο κατάστημα ή τη θέση της κατηγορίας στην λίστα αγοράς του πελάτη (αν χρησιμοποιεί). Βεβαίως, η χωρική θέση των προϊόντων στο κατάστημα, καθώς και η διάταξη του καταστήματος, επηρεάζουν τους πελάτες ως προς τις επιλογές των προϊόντων που θα κάνουν.

Abandonment Rate (Conversion Rate)

Είναι ο δείκτης εμφάνισης μιας κατηγορίας στις επισκέψεις των καταναλωτών, προς τον αριθμό εμφάνισης της ίδιας κατηγορίας στο τελικό καλάθι του καταναλωτή. Σε συνδυασμό με τα δεδομένα από POS, δηλαδή τις αγορές (το καλάθι) που έκανε στη συγκεκριμένη επίσκεψη ο καταναλωτής, μπορεί να βρεθεί το ποσοστό εγκατάλειψης / μη-αγοράς για μια κατηγορία προϊόντων. Κάτι που ήταν δυνατόν να υπολογιστεί μέχρι τώρα μόνο σε ηλεκτρονικά καταστήματα (e-shops) και όχι σε φυσικά (brick and mortar). Αποτελεί μια καλή ένδειξη για την αγοραστική απόδοση της κάθε κατηγορίας προϊόντων. Επιπλέον θα παρατηρηθεί η χρονική διακύμανση του δείκτη στο χρόνο (π.χ. μήνα ή τρίμηνο κλπ.). Για παράδειγμα θα μελετηθεί το αποτέλεσμα των προωθητικών ενεργειών (π.χ. αλλαγών στην τιμή).

Second per Dollar

Οι πωλήσεις αντιπροσωπεύουν τη σημαντικότερη μεταβλητή για τους λιανέμπορους. Ιδιαίτερο ενδιαφέρον παρουσιάζει η σύγκρισή τους με το χρόνο αγορών. Συγκεκριμένα είναι πολύ χρήσιμο μια επιχείρηση να γνωρίζει όχι μόνο τις πωλήσεις ανά επίσκεψη, αλλά και το χρόνο ανά νομισματική μονάδα. Σύμφωνα με τον Sorenson, όσο λιγότερο χρόνο σπαταλάει ο πελάτης, τόσο περισσότερες πωλήσεις θα γίνονται, δηλαδή, οι πελάτες που ξοδεύουν χρήματα πιο γρήγορα, οδηγούν σε μεγαλύτερες συνολικές πωλήσεις (Herb Sorenson, 2009). Για το λόγο αυτό, η μετρική «Second per Dollar» που εισάγει ο Sorenson στο λιανικό εμπόριο, είναι ένας ακόμα σημαντικός δείκτης απόδοσης του καταστήματος.

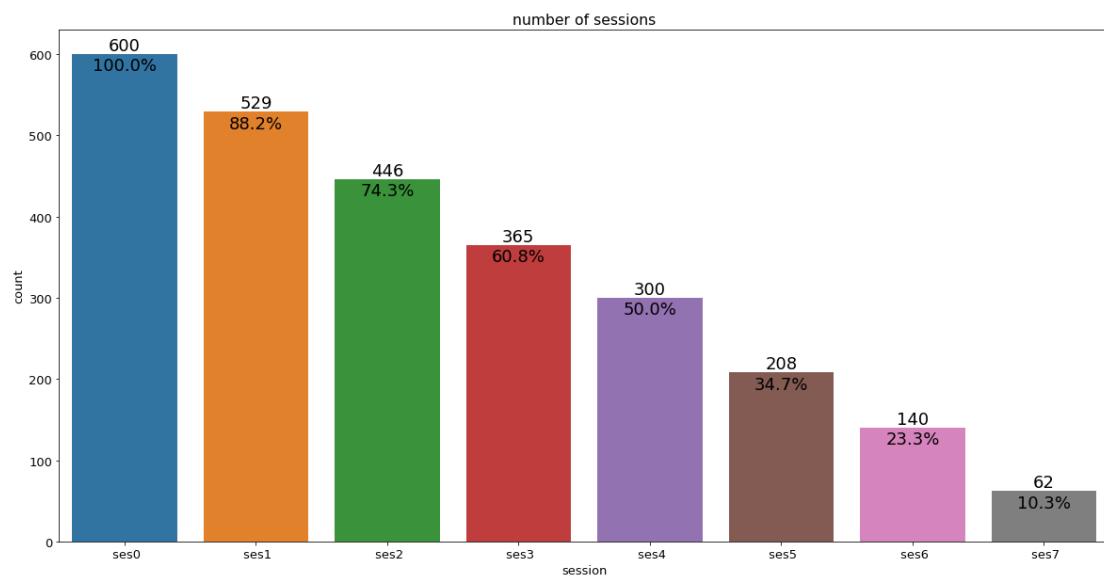
Web metric	In-Store metric
Session/First session	Store visit/First store visit
Unique/Repeat visitor	Unique/Repeat customer
Session duration	Store exit time – Store enter time
Page views per session	Zones visited per store visit
Avg. new/repeat session length	Avg. time spent in store; new/repeat
Avg. time per page	Avg. time per zone
Traffic volume	Total visitors to store
Bounce rate	Ratio of short visits/no visited zones

Πίνακας 8: Αντιστοιχία μετρικών σε ηλεκτρονικό κατάστημα με αυτών σε ένα φυσικό κατάστημα (Yaeli et al., 2014)

5.1.2 Οπτικοποίηση Των Key Performance Indicators (KPIs)

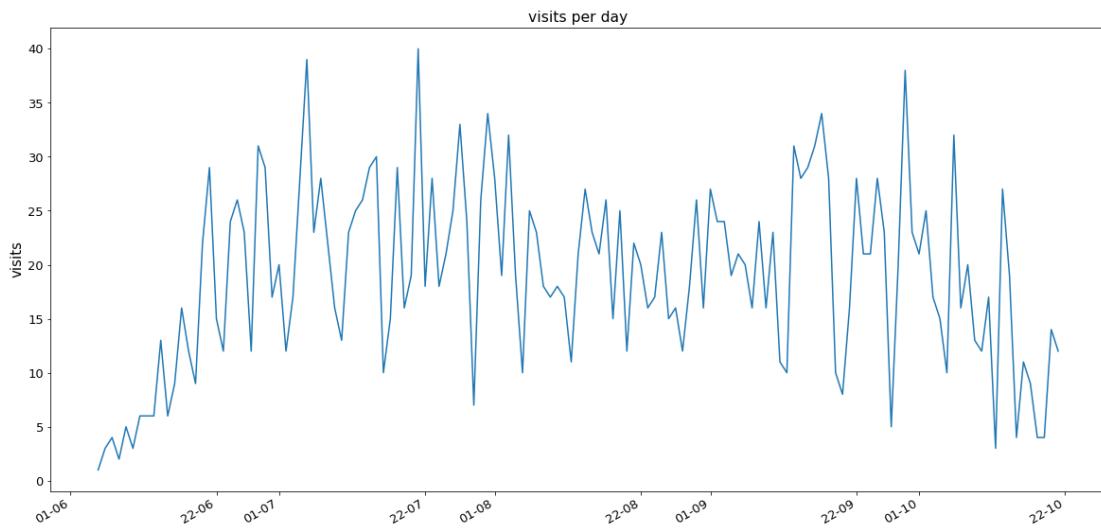
Ο όρος οπτικοποίηση δεδομένων αναφέρεται στην αναπαράσταση δεδομένων χρησιμοποιώντας γραφικά, δισδιάστατες ή και τρισδιάστατες απεικονίσεις των δεδομένων. Τα δεδομένα εμφανίζονται με ένα διαφορετικό τρόπο και η οπτικοποίηση αποκαλύπτει νέες διασυνδέσεις και έννοιες και βοηθά να γίνει η πληροφορία πιο κατανοητή και να καταλήξει ο αναλυτής σε συμπεράσματα για τα δεδομένα.

Στη συνέχεια παρουσιάζονται οι οπτικοποίησεις των δεικτών KPIs που αφορούν στο φυσικό κατάστημα της παρούσας εργασίας.



Εικόνα 33: αριθμός των επισκέψεων των πελάτων

Στο παραπάνω γράφημα, μπορούμε να δούμε πόσοι πελάτες χρησιμοποίησαν την εφαρμογή κατά την επίσκεψή τους. Από τους συνολικά 600 πελάτες, οι 529 (88%) την ξαναχρησιμοποίησαν για δεύτερη φορά. Ενώ οι μισοί πελάτες την χρησιμοποίησαν πέντε φορές και το 10% (62 άτομα) των πελατών, την χρησιμοποίησαν 8 φορές σε αυτό το διάστημα των 4 μηνών.



Εικόνα 34: Γράφημα χρονοδιαγραμμάτων των επισκέψεων σε επίπεδο ημέρας

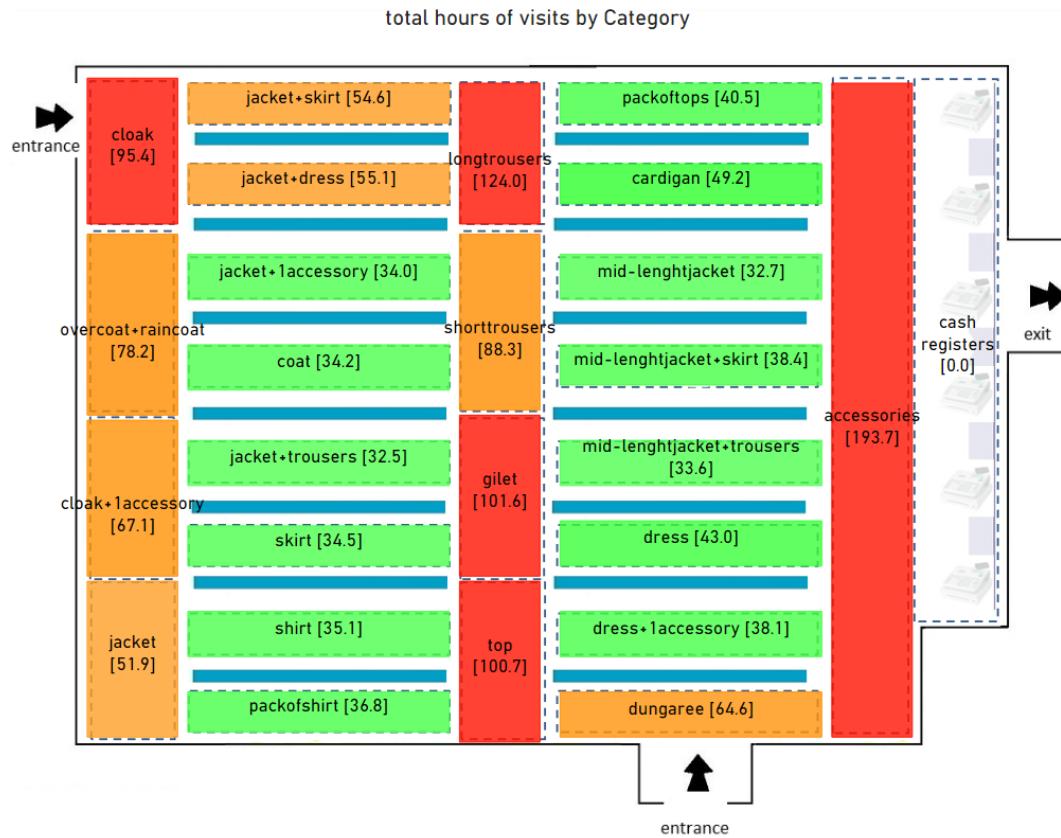
Το παραπάνω γράφημα απεικονίζει την επισκεψιμότητα σε επίπεδο ημέρας. Παρατηρείται ότι η επισκεψιμότητα παρουσιάζει σημαντικές αυξομειώσεις και φαίνεται να επιδεικνύει αυξητική τάση μόνο στην αρχή της περιόδου. Με το γράφημα αυτό, μπορεί να αξιολογηθεί η τάση της χρήσης της εφαρμογής από τους πελάτες.

Μια άλλη πληροφορία που μπορεί να εξαχθεί από αυτά τα δεδομένα, είναι να συλλεχτεί η ώρα επίσκεψης των πελατών στο κατάστημα(Store enter time). Σε μια πρώτη ανάλυση, μπορεί να οπτικοποιηθεί η συχνότητα και να βρεθούν οι ώρες με την μεγαλύτερη επισκεψιμότητα (traffic). Στην παρακάτω εικόνα, παρατηρείται ότι η πιο πιθανή ώρα επίσκεψης ήταν η 12 και 14 το μεσημέρι και ακολουθούν οι 8 και 10 το πρωί. Οπότε μπορεί να συμπεραθεί ότι το κατάστημα έχει μεγαλύτερη επισκεψιμότητα νωρίς το μεσημέρι και το πρωί.



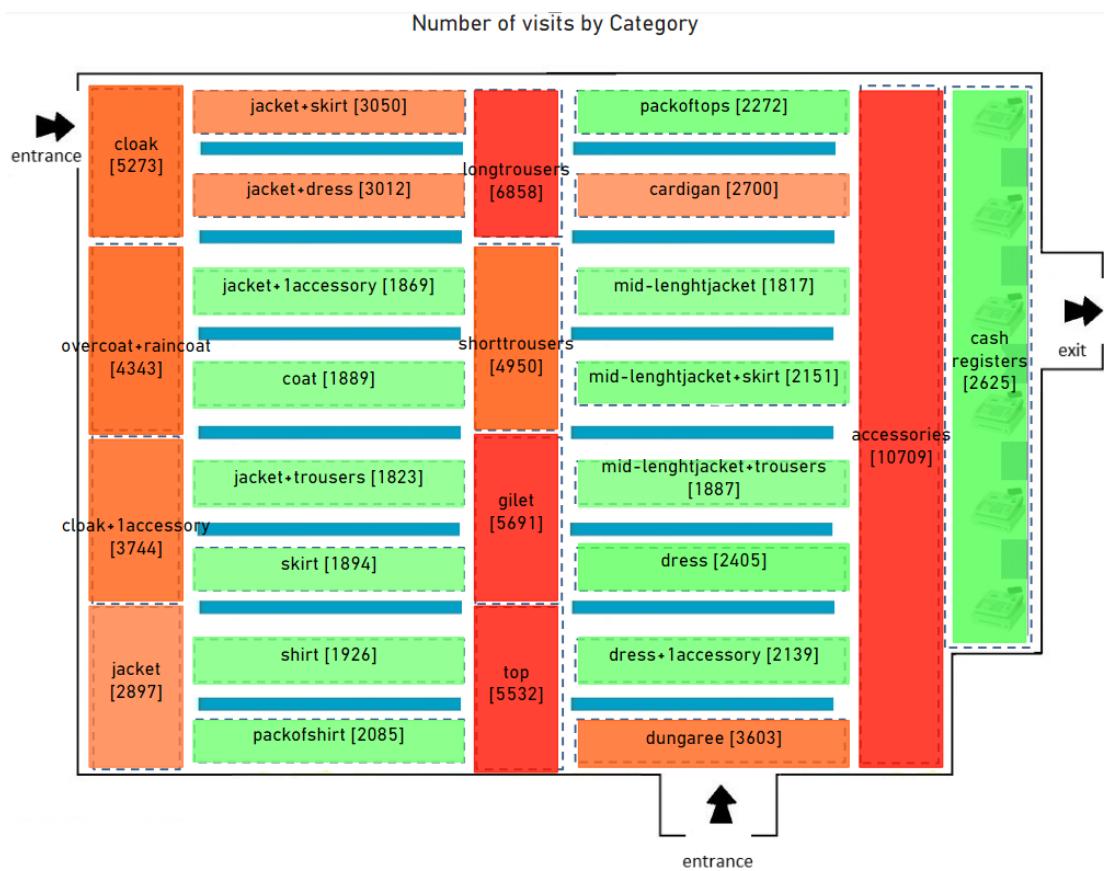
Εικόνα 35: Ιστόγραμμά με τις διάφορες τιμές της ώρα προσέλευσης στο κατάστημα

Παρακάτω απεικονίζονται δείκτες ανά κατηγορία συνδυασμένοι με την χωρική θέση των κατηγοριών που αφορούν.



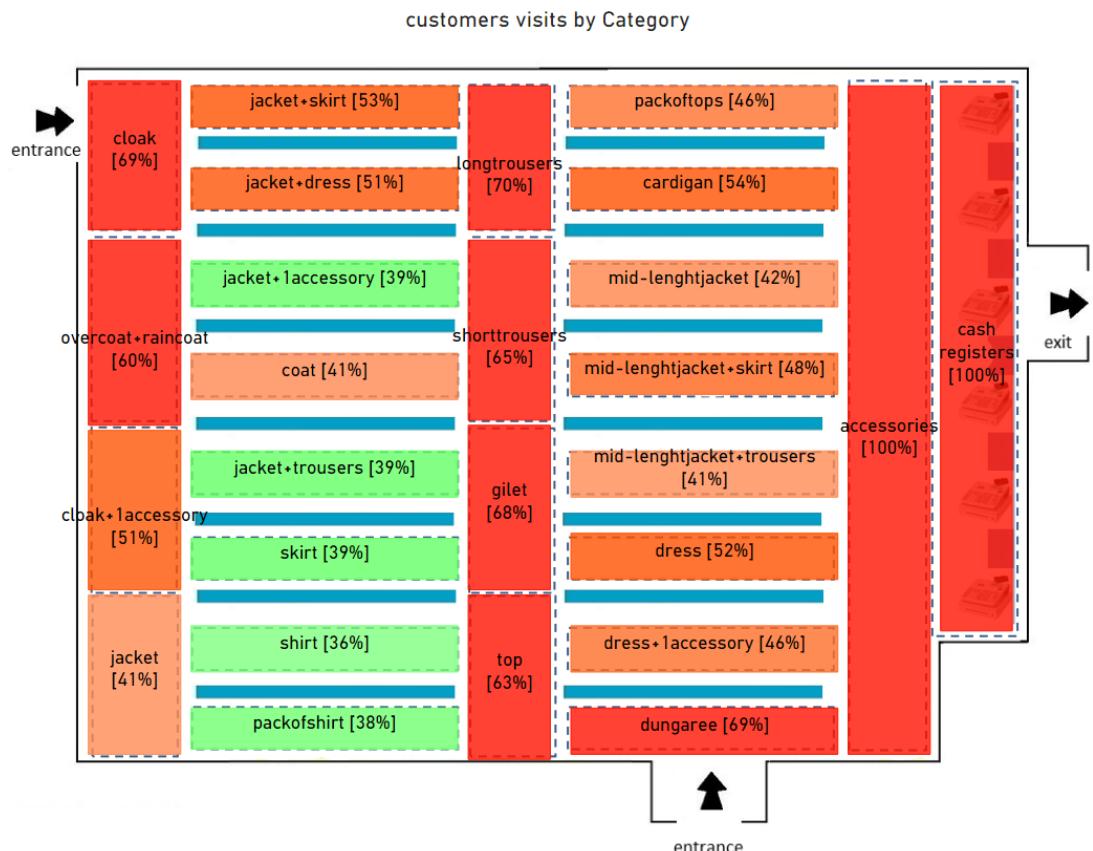
Εικόνα 36: Χρονική διάρκεια των επισκέψεων των πελατών ανά κατηγορία εκφρασμένη σε ώρες

Στην παραπάνω εικόνα απεικονίζονται οι συνολικές ώρες που έχουν περάσει οι πελάτες μέσα στο κατάστημα, μέσα στο χρονικό διάστημα της μελέτης. Αυτή είναι μια πολύ χρήσιμη πληροφορία για τους μάνατζερς του καταστήματος, γιατί τους βοηθάει να κάνουν καλύτερη κατανομή των υπαλλήλων τους στις περιοχές και να ενισχύσουν με περισσότερους υπαλλήλους εκείνες τις περιοχές που παρουσιάζουν μεγαλύτερο χρόνο παραμονής των πελατών. Αυτή η ενέργεια με τη σειρά της, θα τους βοηθήσει να ανεβάσουν την ποιότητα της παρεχόμενης εξυπηρέτησης πελάτων του καταστήματος.



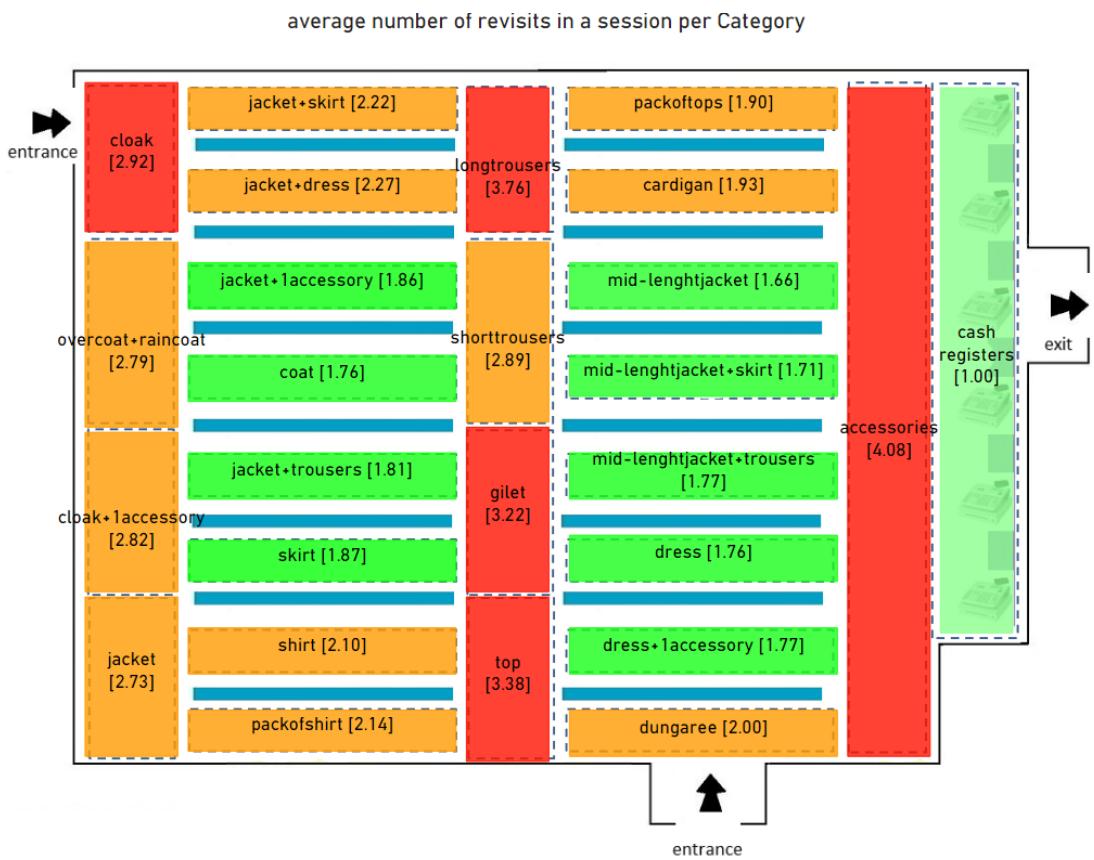
Εικόνα 37: Ολικός αριθμός επισκέψεων ανά κατηγορία από τους πελάτες

Στην παραπάνω εικόνα, φαίνεται ο συνολικός αριθμός επισκέψεων ανά κατηγορία. Δηλαδή πόσες φορές πέρασε ο πελάτης από μια συγκεκριμένη περιοχή. Οι πελάτες μπορεί να βρεθούν σε μια περιοχή παραπάνω από μια φορά κατά τη διάρκεια της επίσκεψής τους στο κατάστημα. Αυτήν την πληροφορία μπορούν να την χρησιμοποιήσουν οι λιανέμποροι, για να βρουν τις κατάλληλες περιοχές για τις προωθητικές τους ενέργειες εντός του καταστήματος. Για παράδειγμα μια τέτοια κατάλληλη περιοχή, είναι η περιοχή με την κατηγορία προϊόντων accessories, όπου μπορούν να τοποθετηθούν διαφημίσεις προϊόντων στις άκρες των διάδρομων του τέμνονται με αυτή την περιοχή (endcaps).



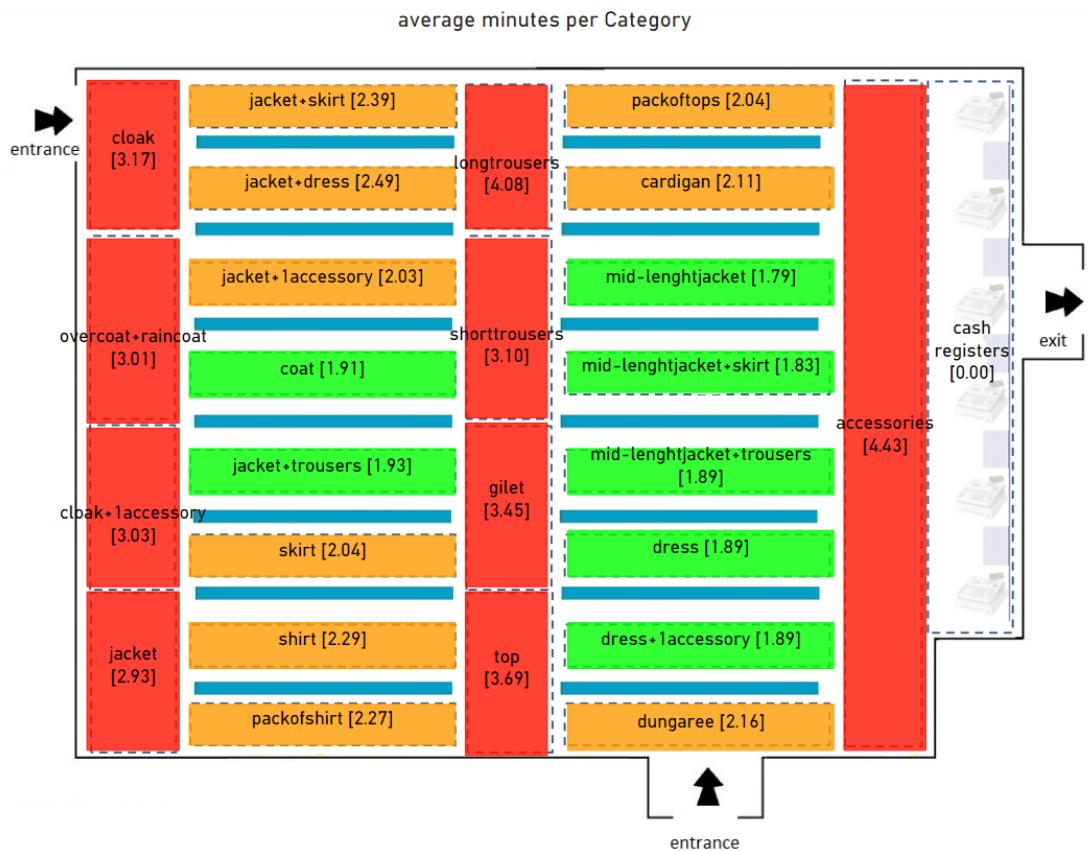
Εικόνα 38: Επισκεψιμότητα ανά κατηγορία

Στην παραπάνω εικόνα φαίνεται σε τι ποσοστό των συνολικών επισκέψεων (sessions) εμφανίζεται μια κατηγορία προϊόντος. Ο παραπάνω δείκτης είναι χρήσιμος στη σύγκριση των περιοχών και στην ανάδειξη των πιο συχνών περιοχών επίσκεψης. Με αυτήν την οπτικοποίηση της μετρικής, γίνεται εύκολο να παρατηρηθούν οι περιοχές υψηλής και χαμηλής επισκεψιμότητας. Συγκεκριμένα παρατηρείται ότι οι πελάτες δεν προτιμούν τις κατηγορίες προϊόντων στην κάτω αριστερή πλευρά του καταστήματος (είναι περιοχές μικρής επισκεψιμότητας), κάτι που μπορεί να μην οφείλεται απαραίτητα στα προϊόντα αλλά στην χωρική τους θέση μέσα στο κατάστημα (ποσό εμφανής είναι η θέση τους και πόσο εύκολο είναι να μετάβουν οι πελάτες σε αυτές).



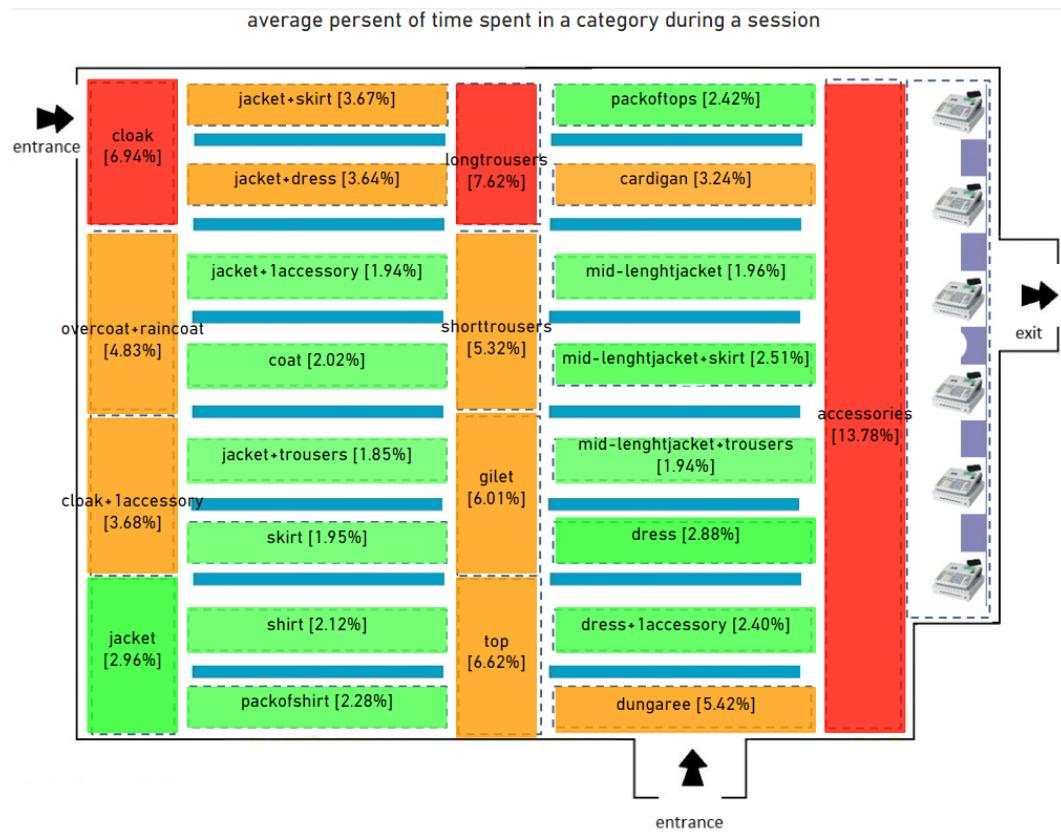
Εικόνα 39: Μέσος αριθμός επανα-επισκέψεων ανά κατηγορία σε μια επίσκεψη

Ο αριθμός επανα-επισκέψης δηλώνει πόσες φορές πέρασε από μια συγκεκριμένη περιοχή ο κάθε πελάτης. Υπολογίστηκε η μέση τιμή αυτών, όπως φαίνεται στην παραπάνω εικόνα. Για παράδειγμα, η περιοχή accessories παρουσιάζει το μεγαλύτερο μέσο όρο επανα-επισκεψιμότητας, καθώς οι πελάτες την επισκέφτηκαν κατά μέσο όρο 4 φορές κατά την παραμονή τους στο κατάστημα. Επίσης παρατηρείται μεγάλος μέσος όρος στους κάθετους διαδρόμους, κάτι που δεν παρατηρείται στους οριζόντιους. Αυτό μπορεί να ερμηνευτεί ως ότι οι πελάτες χρησιμοποιούν τους κάθετους διαδρόμους για να μετακινούνται μεταξύ των περιοχών του καταστήματος.



Εικόνα 40: Μέση διάρκεια επίσκεψης των πελατών ανά κατηγορία εκφρασμένη σε λεπτά

Στην παραπάνω εικόνα παρουσιάζεται ο μέσος χρόνος που παραμένουν οι πελάτες στην περιοχή. Από τη διάταξη της εικόνας φαίνεται ότι οι κάθετοι διάδρομοι (aisles) έχουν μεγαλύτερο χρόνο παραμονής σε σύγκριση με τους οριζόντιους διαδρόμους. Αυτό το συμπέρασμα, σε συνδυασμό με την παρατήρηση της Εικόνας 37, του ολικού αριθμού επισκέψεων, φανερώνει ότι αυτές οι περιοχές αποτελούν τις πιο πολυσύχναστες περιοχές του καταστήματος.



Εικόνα 41: Μέσο ποσοστό του χρόνου που δαπάνησαν οι πελάτες ανά επίσκεψη στην αντίστοιχη κατηγορία

Η παραπάνω εικόνα απεικονίζει το μέσο όρο ανά περιοχή των ποσοστών του χρόνου που δαπάνησαν οι πελάτες στην αντίστοιχη κατηγορία προϊόντων, δια του συνολικού χρόνου παραμονής τους στο κατάστημα. Γίνεται άμεσα αντιληπτό ποιες περιοχές ήταν οι πιο σημαντικές για τους πελάτες, γιατί αυτές οι περιοχές παρουσιάζουν το μεγαλύτερο μέσο ποσοστό. Ωστόσο, αυτό το ποσοστό χρήζει περισσότερης διερεύνησης για την μελέτη της αποδοτικότητας της κάθε περιοχής, γιατί δεν είναι βέβαιο για ποιο λόγο οι πελάτες δαπάνησαν το μεγαλύτερο κομμάτι του χρόνου τους σε αυτήν την περιοχή. Μπορεί να έψαχναν αυτό που ήθελαν να αγοράσουν αλλά δεν το έβρισκαν ή μπορεί να είχαν πραγματικά ενδιαφέρον για τα προϊόντα που φιλοξενεί αυτή η περιοχή.



Εικόνα 42: Σύγκριση των περιοχών μέσω του δείκτη GRP

Για την ανάλυση των δεδομένων της παρούσας εργασίας, χρησιμοποιήθηκε ο δείκτης GPR. Ο δείκτης GPR υπολογίστηκε από το γινόμενο του μέσου χρόνου επίσκεψης της περιοχής με το συνολικό αριθμό επισκεπτών. Στη συνέχεια ταξινομήθηκαν οι κατηγορίες βάσει του αριθμού των επισκεπτών και υπολογίστηκε η ποσοστιαία διαφορά της κάθε περιοχής (n) με την άμεσα ανώτερή της ($n-1$), για να συγκριθούν οι περιοχές με κοντινή επισκεψιμότητα. Στην παραπάνω εικόνα φαίνεται η ποσοστιαία διαφορά των περιοχών. Οι περιοχές με πράσινο χρώμα, έχουν αρνητική ποσοστιαία διαφορά, που σημαίνει ότι σ' αυτές τις περιοχές οι πελάτες παρέμειναν για παραπάνω χρόνο, σε σύγκριση με άλλες περιοχές παρόμοιας επισκεψιμότητας και αυτό τις κάνει πιο αποδοτικές για το κατάστημα.

5.2 Προσφερόμενες Δυνατότητες Της Τεχνολογίας BLE

Beacons Στο Λιανικό Εμπόριο

Σ' αυτήν την ενότητα παρατίθενται οι δυνατότητες που προσφέρει η τεχνολογία BLE beacons στα καταστήματα λιανικού εμπορίου, που αφορούν τη συγκεκριμένη εργασία, οι οποίες μπορούν να ενοποιήσουν τα φυσικά καταστήματα με τα ηλεκτρονικά στη νέα εποχή του ψηφιακού μάρκετινγκ και του ηλεκτρονικού εμπορίου (“iBeacons: The Future of Consumer Engagement,” 2014). Παράδειγμα εφαρμογής της τεχνολογίας είναι οι πρωθητικές ενέργειες που μπορούν να γίνουν μέσα σ' ένα κατάστημα ή πλησίον αυτού π.χ. στη βιτρίνα.



Εικόνα 43: Εφαρμογές των Beacons πηγή (Minds.com)

i) Στοχευμένες προσφορές

Εφόσον ο χρήστης δώσει την έγκρισή του, επιτρέπεται η λήψη ειδοποίησης εντός του καταστήματος. Δηλαδή του αποστέλλονται προσφορές με βάση την ήδη καταγραφείσα καταναλωτική του συμπεριφορά. Η συμπεριφορά του αναλύεται μέσω βασικών στοιχείων την ώρα, την ημέρα, τη διαδρομή που ακολούθησε, τις περιοχές του καταστήματος που επισκέφθηκε, με ποια σειρά καθώς και το χρόνο που παρέμεινε στην κάθε μια, και πολλά άλλα. Η συγκεκριμένη τεχνολογία αυξάνει τις πωλήσεις καθώς εντοπίζει τον καταναλωτή που επισκέφθηκε το κατάστημα και τις κατηγορίες προϊόντων που παρέμεινε περισσότερη ώρα αλλά τελικά δεν ψώνισε ώστε κατά την επόμενη επίσκεψή του στο συγκεκριμένο κατάστημα

ή ακόμα καλύτερα σε οποιοδήποτε της ίδιας αλυσίδας να του προσφερθεί ένα ηλεκτρονικό εκπτωτικό κουπόνι στο smart phone του, ως μέσο προώθησης του προϊόντος.

ii) Ανίχνευση Αγοραστικής Συμπεριφοράς

Η τεχνολογία beacons επιτρέπει στον υπεύθυνο του καταστήματος / πωλητή να παρακολουθεί την αγοραστική συμπεριφορά του καταναλωτή και να του προσφέρει εξατομικευμένα μηνύματα για την προώθηση προϊόντων. Για παράδειγμα παρατηρώντας το σύστημα beacon ότι ένας υποψήφιος αγοραστής παραμένει μπροστά από ένα ράφι πάνω από 3 λεπτά, επιτρέπει στον έμπορο να του στείλει μηνύματα όπως αξιολογήσεις για το προϊόν ή τη γκάμα επιλογών του (νούμερο, χρώμα κλπ). Ακόμα καλύτερα μπορεί αναλύοντας το ιστορικό αλλαγών, να υπενθυμίσει στον καταναλωτή ότι το συγκεκριμένο χρώμα μπλούζας ταιριάζει με το παντελόνι που αγόρασε πρόσφατα από το ίδιο κατάστημα.

iii) Ενοποίηση Online/Offline Καταναλωτικής Συμπεριφοράς (Omni-Channel)

Η τεχνολογία beacons δίνει τη δυνατότητα σ' ένα πωλητή να συγκεράσει πληροφορίες για την αγοραστική συμπεριφορά του καταναλωτή τόσο στο φυσικό όσο και στο ηλεκτρονικό κατάστημα. Γι' αυτό η άνω συγχώνευση ονομάστηκε από τους ειδικούς στο λιανικό εμπόριο omni-channel. Αυτό ήταν μια μεγάλη εξέλιξη για την προώθηση των πωλήσεων καθώς μέχρι σήμερα στο φυσικό κατάστημα η αγοραστική συμπεριφορά μπορούσε να παρατηρηθεί μόνο στο ταμείο όπου ο καταναλωτής χρησιμοποιούσε την κάρτα του καταστήματος. Ενώ η καλύτερη προώθηση απαιτεί την παρατήρηση του καταναλωτή σε πραγματικό χρόνο κάτι που μέχρι πρόσφατα ήταν δυνατό μόνο στο ηλεκτρονικό κατάστημα. Επιπλέον η τεχνολογία των beacons λειτουργεί ως μέσο υπενθύμισης στον έμπορο του λιανικού συστήματος για την προώθηση στον καταναλωτή προϊόντα που κατά την επίσκεψη του στο ηλεκτρονικό κατάστημα τοποθετήθηκαν στο ηλεκτρονικό καλάθι αλλά τελικά δεν αγοράστηκαν. Λειτουργούν δηλαδή οι ηλεκτρονικές και φυσικές πωλήσεις συμπληρωματικά.

iv) Δημιουργία Συσχέτισης με τον Καταναλωτή

Η τεχνολογία beacon παρακολουθεί την αγοραστική συμπεριφορά του καταναλωτή και του παρέχει γενικότερες πληροφορίες που ταιριάζουν πάντα στο προφίλ του. Δηλαδή παρατηρώντας τα προϊόντα που προτιμά κατανοεί τους τομείς ενδιαφέροντος του και του αποστέλλει πληροφορίες (π.χ. για τα sport που προτιμά, τα χόμπι του κλπ.) δημιουργώντας έτσι μια σχέση συνεργασίας και εμπιστοσύνης μεταξύ εμπόρου και καταναλωτή.



v) Προσέλκυση του Καταναλωτή στο Κατάστημα

Η δυναμική της τεχνολογίας beacon που την κάνει ιδιαίτερη είναι η ικανότητά της να προσελκύει τον καταναλωτή και έξω από το φυσικό κατάστημα. Ένα beacon μπορεί να τοποθετηθεί σε βιτρίνα καταστήματος ή σε στάση λεωφορείου ή σε πινακίδα στο δρόμο ή σε σταθμό του μετρό και να λειτουργήσει ως πομπός διαφημιστικών μηνυμάτων σε περαστικούς που θα τους προσελκύσει στο κατάστημα. Είναι ευνόητο ότι λειτουργεί ως «φάρος» προσέλκυσης σε υποψήφιους αγοραστές.

vi) Αξιολόγηση Προσωπικού

Η τεχνολογία beacon μπορεί να αξιοποιηθεί μέσα στο κατάστημα εκτός από την προώθηση πωλήσεων και στην αξιολόγηση του προσωπικού. Δηλαδή μπορεί να μετρηθεί ο χρόνος που δαπανά ένας εργαζόμενος σε μια διεργασία πχ. Πόσο χρόνο ασχολείται ένας εργαζόμενος του καταστήματος με έναν συγκεκριμένο πελάτη. Με αυτό τα δεδομένα μπορούν τα αξιοποιηθούν από τον εκάστοτε οργανισμό για τη δημιουργία δεικτών απόδοσης που με την σειρά τους θα βοηθήσουν στην καλύτερη εκπαίδευση του προσωπικού.

vii) Έγκαιρος Ανεφοδιασμός (OSA)

Επιπρόσθετα τα beacon αξιοποιούνται και για τον έγκαιρο ανεφοδιασμό του καταστήματος σε επίπεδο ραφιού δίνοντας πληροφορίες για μια περιοχή καταστήματος όπου παρατηρείται αυξανόμενη κινητικότητα καταναλωτών (On-shelf Availability). Σε μια έρευνα της PWC (PWC, 2017) για το λιανεμπόριο επισημαίνεται το μέγεθος του προβλήματος σχετικά με τη διαθεσιμότητα των προϊόντων: οι καταναλωτές σε περίπτωση που δεν βρουν το προϊόν που ψάχνουν, το 49% διαλέγει εναλλακτικό προϊόν ενώ το 39% φεύγει χωρίς να προχωρήσει σε αγορά και το 12% πάει σε κατάστημα ανταγωνιστή.

Από τα ανωτέρω συνάγεται ότι οι ευκαιρίες που προσφέρει η νέα τεχνολογία beacon στα καταστήματα λιανικής πώλησης είναι πολλές και συνεχώς αυξανόμενες, με σημαντικότερες την παροχή πληροφοριών για τον υποψήφιο καταναλωτή σε πραγματικό χρόνο και τη συλλογή δεδομένων της συμπεριφοράς των καταναλωτών.

5.2.1 Παραδείγματα Mobile Marketing Με Τη Χρήση Της Τεχνολογίας BLE

Beacons Στο Λιανικό Εμπόριο

Τα καταστήματα λιανικής πώλησης Walmart με έδρα στην Αμερική έχουν συνεργαστεί με την General Electric στην εγκατάσταση beacons στα καταστήματά τους. Στόχος είναι η ανάπτυξη του mobile marketing και η ενδυνάμωση της πιστότητας των πελατών τους μέσω προωθητικών ενεργειών όπως κουπονιών, την παροχή πληροφοριών για τα προϊόντα και την πλοιόγηση μέσα στα καταστήματα με τη χρήση της εφαρμογής της εταιρείας για κινητά (9to5mac.com, 2014).

To Swan Center είναι το πρώτο εμπορικό κέντρο στην Αγγλία που έχει επιτρέψει την οργάνωση προωθητικών καμπανιών προϊόντων εντός των καταστημάτων μέσω beacons. (informationage.com, 2014)

Τα πολυκαταστήματα House of Fraser στην Αγγλία έχουν τοποθετήσει beacons στις κούκλες στις βιτρίνες των καταστημάτων, με αποτέλεσμα ο περαστικός να ελέγχει από το κινητό του τηλέφωνο μέσω μιας προ εγκαταστημένης εφαρμογής, πληροφορίες για τα ρούχα που φοράει η κούκλα και αν επιθυμεί να το αγοράσει μέσω του ηλεκτρονικού καταστήματος. Επιπλέον οι καταναλωτές φτιάχνοντας ένα προφίλ στην εφαρμογή και βάσει των λιστών επιθυμιών τους και μέσω της καταγραφής του ιστορικού των αγορών τους μπορεί η εφαρμογή να στέλνει στους χρηστές εξατομικευμένες ειδοποιήσεις αν έχουν ανοιχτή την κεραία του Bluetooth της κινητής συσκευής τους και βρεθούν σε απόσταση μικρότερη των 50 μέτρων από ένα κατάστημα. Μ' αυτό τον τρόπο ενοποιούν το φυσικό με το ηλεκτρονικό κατάστημα τους μέσω του m-commerce. (rettaidive.com)

Ο εμπορικός δρόμος του Λονδίνου, Regent Street, είναι γνωστός για την πληθώρα των καταστημάτων και εστιατορίων που υπάρχουν σε αυτόν. Ο δρόμος αυτός έδωσε το όνομα του και σε μια εφαρμογή γνωστή ως Regent Street app η οποία επιτρέπει στους καταναλωτές να λαμβάνουν μηνύματα καθώς στις εισόδους των συνεργαζόμενων καταστημάτων έχουν τοποθετηθεί beacons τα οποία ανέρχονται σε πάνω από 130. Οι καταστηματάρχες δηλώνουν πως έχουν δει τις εξαργυρώσεις προσφορών μέσω της εφαρμογής για κινητά να αυξάνονται κατά δέκα φορές και το ποσοστό ανταπόκρισης των καταναλωτών στις ειδοποιήσεις που δέχτηκαν να είναι στο 35% (Beaconstac.com, 2019).

Λοιποί Χώροι Εφαρμογής της Τεχνολογίας BLE Beacons

Οι δυνατότητες που προσφέρει η τεχνολογία Beacon επεκτείνονται και σ' άλλους τομείς όπως: μουσεία, χώροι υποδοχής, εκπαιδευτικά ιδρύματα, αθλητικά στάδια, φορείς υγειονομικής περίθαλψης, χώροι εστίασης, καζίνο, χώροι στάθμευσης αυτοκινήτων, μέσα μαζικής μεταφοράς. Παρακάτω γίνεται μια παρουσίαση των πιθανών σεναρίων εφαρμογής της τεχνολογίας beacon ανά τομέα:

5.3 Αποδοχή Των Νέων Τεχνολογιών Από Τους Χρήστες

Ωστόσο υφίστανται εμπόδια στην ανάπτυξη της εγγύτητας ως μέσο προώθησης λόγω της καχυποψίας κάποιων καταναλωτών στην επερχόμενη υιοθέτηση των υπηρεσιών προώθησης βάσει θέσης από τις εταιρείες και τις κυβερνήσεις. Έχει μελετηθεί η πρόθεση χρήσης του αγοραστικού κοινού για την εφαρμογή συστημάτων καθορισμού θέσης και είναι κοινώς αποδεκτό το γεγονός ότι η προστασία της ιδιωτικότητας του καταναλωτή είναι σημαντική σε ένα κατάστημα λιανικής. Για παράδειγμα σε έρευνες που διεξήχθησαν σε focus groups στην Γερμανία (σε πελάτες καταστημάτων ηλεκτρονικών με στρωματοποιημένη δειγματοληψία) και στην Αγγλία (σε φοιτητές που στην πλειοψηφία τους είχαν smart phones και ήξεραν τις εφαρμογές του Bluetooth), οι ερευνητές κατέληξαν στο συμπέρασμα ότι οι χρήστες ήταν απρόθυμοι στη χρήση τέτοιων τεχνολογιών γιατί φοβόντουσαν ότι παραβίαζαν προσωπικά δεδομένα ανεξάρτητα της εξοικείωσής τους με την τεχνολογία (Cockrill, Goode, & White, 2011) (Müller-Seitz, Dautzenberg, Creusen, 2009) και το πόσο καλά ήταν πληροφορημένοι για τα οφέλη αυτών των τεχνολογιών (Günther & Spiekermann, 2007). Σε παρόμοια αποτελέσματα κατέληξαν και οι Πραματάρη και Θεοτόκης σε έρευνά τους, για την αποδοχή από το ελληνικό αγοραστικό κοινό των παρεχόμενων υπηρεσιών/προσφορών με χρήση RFID στη λιανική. Αναφέρουν το άγχος για τη χρήση νέων τεχνολογιών και τη συλλογή των προσωπικών τους δεδομένων (προστασία της ιδιωτικότητάς τους) ως σημαντικό αποτρεπτικό λόγο για την αποδοχή και χρήση αυτών των υπηρεσιών από τους καταναλωτές (Pramatari & Theotokis, 2011).

Η νομοθεσία περί προστασίας προσωπικών δεδομένων (GDPR) και οι ρυθμίσεις των κινητών τηλεφώνων και των ανάλογων εφαρμογών πρέπει να επιτρέπουν στους καταναλωτές κατά την εγγραφή τους να ελέγχουν οι ίδιοι σε ποιες υπηρεσίες συναινούν (J. Wang & Loui, 2009). Συνεπώς προκύπτει μια πρόκληση για τους marketers και τους λιανέμπορους να δημιουργήσουν στους καταναλωτές, μέσω καμπανιών, κίνητρα για τη χρήση (loyalty systems) και εμπιστοσύνη στην τεχνολογία beacon ώστε οι τελευταίοι να εγκαταστήσουν στις συσκευές των κινητών τηλεφώνων τους (smart phone) αντίστοιχες εφαρμογές αυτής της τεχνολογίας και να συμμετέχουν σε αντίστοιχες υπηρεσίες προώθησης, προκειμένου να ωφεληθούν και οι δυο πλευρές.



6

Επίλογος

6.1 Σύνοψη και Συμπεράσματα

Αυτή η εργασία, αναγνωρίζοντας την ανάγκη των λιανεμπορίων να κατανοήσουν καλύτερα το αγοραστικό τους κοινό, προσπαθεί να σχεδιάσει μια μέθοδο ανάλυσης (framework) της κίνησης των πελατών μέσα σε ένα κατάστημα, ώστε να ενισχύσει τη λήψη αποφάσεων της διοίκησης του καταστήματος. Η παρούσα διπλωματική εργασία κάνει χρήση της μεθοδολογίας CRISP-DM.

Κάποια συμπεράσματα, βασισμένα στην επίσημη βιβλιογραφία, είναι τα εξής: Από τότε που το GPS εισήχθη για εμπορική χρήση τη δεκαετία του 90, όλο και περισσότερες υπηρεσίες που βασίζονται στον εντοπισμό θέσης έχουν εμφανιστεί. Το GPS έχει επικρατήσει ως πρότυπο σύστημα εντοπισμού θέσης σε εξωτερικό χώρο. Αντίστοιχο πρότυπο για χρήση σε εσωτερικούς χώρους δεν υπάρχει. Ωστόσο τα τελευταία χρόνια, όλο και περισσότερη έρευνα γίνεται σε εσωτερικά συστήματα παρακολούθησης και έχουν προταθεί για χρήση ραδιοκύματα, ακουστικά σήματα, κάμερες, πληροφορίες από κινητές συσκευές και μαγνητικά πεδία. Για πολλά από αυτά, είναι ακριβή η υλοποίηση της εφαρμογή τους. Μια από αυτές τις υλοποιήσεις είναι τα BLEs, όπου για ένα σύστημα θέσης, είναι εύκολο να χρησιμοποιηθούν με λογικό κόστος και να παρέχουν ικανοποιητική ακρίβεια και εμβέλεια. Επιπλέον όλες οι κινητές συσκευές διαθέτουν κεραία Bluetooth.

Το θέμα της προστασίας προσωπικών δεδομένων είναι σημαντικό, και θα πρέπει να είναι ξεκάθαρο στους καταναλωτές ποια δεδομένα τους χρησιμοποιούνται και ποια είναι η χρήση τους. Επιπλέον, για να πετύχει ένα παρόμοιο σύστημα στη λιανική, είναι απαραίτητο οι καταναλωτές να συμμετέχουν στην διαδικασία (opt-in), παρά να απέχουν (opt-out), για αυτό

το λόγο πρέπει να γίνει κατάλληλη μελέτη πριν από κάθε εφαρμογή, τι αξία μπορεί να έχει για το κατάστημα και τι αποδοχή θα έχει από το αγοραστικό του κοινό.

Από τη μέθοδο ανάλυσης που παρουσιάστηκε, προκύπτει ότι επιτεύχθηκαν οι επιχειρησιακοί στόχοι που είχαν τεθεί, δηλαδή μέσω της κατάτμησης των επισκέψεων των πελατών, κατέστη δυνατό να παρατηρηθούν επισκέψεις με συγκεκριμένο αγοραστικό στόχο (shopping mission). Οι πελάτες αυτών των συστάδων ήξεραν τι ήθελαν να αγοράσουν και αυτό φαίνεται από τα ποσοστά επίσκεψης στις χαρακτηριστικές περιοχές της αντίστοιχης συστάδας.

Υπάρχουν συστάδες που δημιουργήθηκαν από τον αλγόριθμο, ακολουθώντας τους χωρικούς περιορισμούς του καταστήματος και σε συνδυασμό με το χαρακτηριστικό ότι οι επισκέψεις είχαν μικρά μονοπάτια, εξάγεται το συμπέρασμα ότι οι πελάτες που μπήκαν από την ίδια είσοδο έκαναν παρόμοιες κλειστές διαδρομές. Μια παραπάνω υπόθεση είναι ότι η είσοδος του καταστήματος που μπήκε ο πελάτης, καθόρισε τη διαδρομή του.

Αυτή η μέθοδος ανάλυσης δεδομένων, μπορεί να βοηθήσει τους λιανέμπορους να οργανώσουν και να σχεδιάσουν κατάλληλα τις προωθητικές και διαφημιστικές τους ενέργειες (προώθηση, προσφορές) εντός και εκτός του καταστήματος, για να έχουν τα μέγιστα αποτελέσματα, δηλαδή την αύξηση των αυθόρυμητων αγορών και των παράλληλων πωλήσεων σε συγκεκριμένα προϊόντα, τα οποία θα αποφασίσουν πότε (ποιες μέρες) και πού θα τοποθετηθούν (σημεία ενδιαφέροντος) αυτά.

Για το λόγο αυτό, οι λιανέμποροι, καλό είναι να συμπεριλάβουν στη μελέτη των δράσεων τους την επισκεψιμότητα ανά ημέρα και ανά περιοχή του καταστήματος.

6.2 Μελλοντικές Επεκτάσεις

Η παρούσα εργασία αποτελεί μια προκαταρκτική μελέτη επιχειρηματικής αναλυτικής για την ανάλυση χωροχρονικών δεδομένων των χρηστών. Αν κάποια εταιρία λιανικής θελήσει να αξιοποιήσει χωροχρονικά δεδομένα, αυτή η διπλωματική μπορεί να δώσει κάποιες κατευθύνσεις σχετικά με τις δυνατότητες που μπορεί να της παρέχει και τους στόχους που μπορεί να θέσει το κατάστημα βάσει των μετρικών που παρουσιάστηκαν.

Ως επέκταση της παρούσας εργασίας, προτείνεται να σχεδιασθεί και να υλοποιηθεί μια εφαρμογή για έξυπνες κινητές συσκευές η οποία θα βασίζεται στο σύστημα συλλογής δεδομένων, που αναφέρθηκε στο κεφάλαιο 4, για την εύρεση θέσης των καταναλωτών. Τα δεδομένα που θα συλλέξει η εφαρμογή, σε συνδυασμό με τα δεδομένα αγορών αυτών των πελατών που θα έχει το κατάστημα, θα αναλυθούν με σκοπό να μελετηθεί η συσχέτιση της κίνησης των πελάτων με τις αγορές τους.

Αυτή η τεχνολογία και οι εφαρμογές της μπορούν να οδηγήσουν τους λιανέμπορους στο χώρο της εξατομικευμένης πληροφόρησης και υπηρεσιών μέσα σε ένα φυσικό κατάστημα, όπως εφαρμόζονται και στα ηλεκτρονικά καταστήματα. Μια χρήσιμη προέκτασή της θα είναι να βοηθάει τους πελάτες τους καταστήματος που χρησιμοποιούν την εφαρμογή, ώστε να πλοηγηθούν στην κατηγορία των προϊόντων που αναζητούν, δηλαδή ως ένας ψηφιακός πλοηγός του καταστήματος.

Επιπρόσθετα, μια επέκταση της εφαρμογής είναι η δημιουργία ενός συστήματος συστάσεων (recommendation system) από τα δεδομένα πελατών που έχουν συλλεχθεί. Αφού καταγραφεί πόσο συχνά επισκέπτεται ένας πελάτης μια συγκεκριμένη κατηγορία, αν έχει περάσει καιρός που δεν την έχει επισκεφθεί, να παρακινείται από το σύστημα να την επανεπισκεφτεί. Επίσης να χρησιμοποιηθεί ο χρόνος που πέρασαν οι χρήστες στην κατηγορία, ως μετρική για ένα σύστημα συστάσεων, βασισμένο στο συνεργατικό φιλτράρισμα (collaborating filtering). Δηλαδή να αναλυθούν τα δεδομένα, με στόχο να προσδιοριστούν οι χρήστες με παρόμοιες προτιμήσεις κατηγοριών, που περνάνε πολύ χρόνο σε μια κατηγορία και να καθοριστεί ο βαθμός ομοιότητάς τους, και, αν ένας συγκεκριμένος πελάτης δεν έχει ποτέ επισκεφτεί μια κατηγορία, αλλά άλλοι πελάτες με παρόμοιες προτιμήσεις με αυτόν έχουν, να του προτείνει την αντίστοιχη κατηγορία, γιατί πιθανότατα να έχει ενδιαφέρον και για τον ίδιο.

Τέλος, μπορούν να γίνουν έρευνες ερωτηματολογίου σε δείγματα του αγοραστικού κοινού, ώστε να μελετηθούν οι ανάγκες των καταναλωτών, δηλαδή τι χαρακτηριστικά (features) θέλουν να περιλαμβάνει η εφαρμογή. Στη συνέχεια, θα πρέπει να πραγματοποιηθεί και μελέτη σε ομάδες ενδιαφέροντος (focus groups), ώστε να ελεγχθεί η ευχρηστία και αποτελεσματικότητα της εφαρμογής.

Βιβλιογραφία

- 49's_Stadium, & Meridian. (2014). How The 49ers Are Using Beacons To Help You Find Hot Dogs And Beer. Retrieved from <https://techcrunch.com/2014/11/04/how-the-49ers-are-using-beacons-to-help-you-find-hot-dogs-and-beer/>
- 9to5mac. (2014). GE integrates iBeacons in new LED lighting fixtures rolling out in Walmart & other retailers. Retrieved from <https://9to5mac.com/2014/05/29/ge-integrates-iota-beacons-in-new-led-lighting-fixtures-rolling-out-in-walmart-other-retailers/>
- Accenture. (2015). National Gallery Singapore and Accenture Launch the Gallery Explorer Mobile App. Retrieved May 1, 2020, from https://newsroom.accenture.com/news/national-gallery-singapore-and-accenture-launch-the-gallery-explorer-mobile-app.htm?utm_content=buffer2ba42&utm_medium=social&utm_source=twitter.com&utm_campaign=buffer
- Aggarwal, C. C., & Reddy, C. K. (2014). *Data clustering : algorithms and applications*. CRC Press.
- Al Delail, B., Weruaga, L., Zemerly, J., & Ng, J. (2013). Indoor localization and navigation using smartphones augmented reality and inertial tracking. In *Proceedings of the IEEE International Conference on Electronics, Circuits, and Systems* (pp. 929–932). <https://doi.org/10.1109/ICECS.2013.6815564>
- ArcNews. (2012). Location Analytics: The Next Big Step in Business Analysis. Retrieved June 1, 2020, from <https://www.esri.com/news/arcnews/fall12articles/location-analytics-the-next-big-step-in-business-analysis.html>
- Beaconstac.com. (2019). The Regent Street beacon app yields in-store results and more. Retrieved from <https://blog.beaconstac.com/2016/08/best-of-beacons-this-week-the-regent-street-beacon-app-yields-in-store-results-and-more/>
- Brena, R. F., García-Vázquez, J. P., Galván-Tejada, C. E., Muñoz-Rodríguez, D., Vargas-Rosales, C., & Fangmeyer, J. (2017). Evolution of Indoor Positioning Technologies: A Survey. *Journal of Sensors*, 2017, 2630413. <https://doi.org/10.1155/2017/2630413>
- Bruner, G. C., & Kumar, A. (2007). Attitude toward Location-based Advertising. *Journal of Interactive Advertising*, 7(2), 3–15. <https://doi.org/10.1080/15252019.2007.10722127>
- Business, & Insider. (2016). Beacons are thriving in sports arenas. Retrieved from



- https://www.businessinsider.com/beacons-are-thriving-in-sports-arenas-2016-8
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. R. H., & Wirth, R. (2000). CRISP-DM 1.0: Step-by-step data mining guide.
- Cockrill, A., Goode, M., & White, A. (2011). Bluetooth Proximity Marketing: Potential and Barriers. *Journal of Advertising Research*, 51, 298–312.
- Ferracuti, N., Norscini, C., Frontoni, E., Gabellini, P., Paolanti, M., & Placidi, V. (2019). A business application of RTLS technology in Intelligent Retail Environment: Defining the shopper's preferred path and its segmentation. *Journal of Retailing and Consumer Services*, 47, 184–194. https://doi.org/10.1016/j.jretconser.2018.11.005
- Fujino, T., Kitazawa, M., Yamada, T., Takahashi, M., Yamamoto, G., Yoshikawa, A., & Terano, T. (2014). Analyzing In-store Shopping Paths from Indirect Observation with RFID Tags Communication Data. *Journal on Innovation and Sustainability. RISUS ISSN 2179-3565*, 5, 88–96. https://doi.org/10.24212/2179-3565.2014v5i1p88-96
- Gomez, C., Oller, J., & Paradells, J. (2012). Overview and evaluation of bluetooth low energy: An emerging low-power wireless technology. *Sensors (Switzerland)*, 12(9), 11734–11753. https://doi.org/10.3390/s120911734
- grandviewresearch.com. (n.d.). Internet of Things In Retail Market Size, Trends Report, 2018–2025. Retrieved May 12, 2020, from https://www.grandviewresearch.com/industry-analysis/internet-of-things-iot-retail-market
- Griva, A., Bardaki, C., Pramatari, K., & Papakiriakopoulos, D. (2018). Retail business analytics: Customer visit segmentation using market basket data. *Expert Systems with Applications*, 100, 1–16. https://doi.org/10.1016/j.eswa.2018.01.029
- Günther, O., & Spiekermann, S. (2007). RFID and the Perception of Control: The Consumer's View. *Engineering Management Review, IEEE*, 48, 40. https://doi.org/10.1109/EMR.2007.382635
- Huang, Z. (1998). Extensions to the k-Means Algorithm for Clustering Large Data Sets with Categorical Values. *Data Mining and Knowledge Discovery*, 2(3), 283–304. https://doi.org/10.1023/A:1009769707641
- Huang, Z., & Huang, Z. (1997). Clustering large data sets with mixed numeric and categorical values. *IN THE FIRST PACIFIC-ASIA CONFERENCE ON KNOWLEDGE DISCOVERY AND DATA MINING*, 21–34. Retrieved from http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.94.9984
- Hui, S. K., Fader, P. S., & Bradlow, E. T. (2009). The traveling salesman goes shopping: The systematic deviations of grocery paths from TSP optimality. *Marketing Science*, 28(3), 566–572. https://doi.org/10.1287/mksc.1080.0402



- iBeacons: The Future of Consumer Engagement. (2014). Retrieved from <https://www.visualistan.com/2014/07/ibeacons-future-of-consumer-engagement.html>
- informationage.com. (2014). First UK shopping centre implements Beacon technology. Retrieved March 5, 2020, from <https://www.information-age.com/first-uk-shopping-centre-implements-beacon-technology-123457806/>
- Jung, I., Kwon, Y. S., & Lee, Y. (2012). A Sequence Pattern Matching Approach to Shopping Path Clustering. *International Conference on Industrial Engineering and Operations Management ISTambul*, 2364–2372.
- Kalgotra, P., & Sharda, R. (2016). Progression analysis of signals: Extending CRISP-DM to stream analytics. <https://doi.org/10.1109/BigData.2016.7840937>
- Kaneko, Y., & Yada, K. (2016). Fractal Dimension of Shopping Path: Influence on Purchase Behavior in a Supermarket. *Procedia Computer Science*, 96(September), 1764–1771. <https://doi.org/10.1016/j.procs.2016.08.225>
- Karlis, D. (2005). *Multivariate statistical analysis* (Greek). Athens: Stamoulis Publications.
- KDnuggets. (2007a). Poll: CRISP-DM the Most Popular DM Methodology.
- KDnuggets. (2007b). Poll: Data Mining Methodology.
- Kodinariya, T., & Makwana, P. R. (2013). Review on Determining of Cluster in K-means Clustering. *International Journal of Advance Research in Computer Science and Management Studies*, 1, 90–95.
- Kumar, L., Singh, H., & Kaur, R. (2012). Web Analytics and Metrics: A Survey. In *Proceedings of the International Conference on Advances in Computing, Communications and Informatics* (pp. 966–971). New York, NY, USA: Association for Computing Machinery. <https://doi.org/10.1145/2345396.2345552>
- Larson, J. S., Bradlow, E. T., Fader, P. S., Larson, J. S., Bradlow, E. T., & Fader, P. S. (2005). An Exploratory Look at Supermarket Shopping Paths. *International Journal of Research in Marketing*, 22(4), 395–414. <https://doi.org/10.1016/j.ijresmar.2005.09.005>
- Levesque, N., Boeck, H., Durif, F., & Bilolo, A. (2015). *The Impact of Proximity Marketing on Consumer Reaction and Firm Performance The Impact of Proximity Marketing on Consumer Reaction and Firm Performance: A Conceptual and Integrative Model*. Americas Conference on Information Systems.
- Liebig, T., & Wagoum, A. U. K. (2012). Modelling microscopic pedestrian mobility using bluetooth. *ICAART 2012 - Proceedings of the 4th International Conference on Agents and Artificial Intelligence*, 2, 270–275.
- Liu, H., Darabi, H., Banerjee, P., & Liu, J. (2007, November). Survey of wireless indoor positioning techniques and systems. *IEEE Transactions on Systems, Man and Cybernetics*.



Part C: Applications and Reviews. <https://doi.org/10.1109/TSMCC.2007.905750>

- MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 1: Statistics* (pp. 281–297). Berkeley, Calif.: University of California Press. Retrieved from <https://projecteuclid.org/euclid.bsmsp/1200512992>
- Maronna, R. (2016). Charu C. Aggarwal and Chandan K. Reddy (eds.): Data clustering: algorithms and applications. *Statistical Papers*, 57(2), 565–566. <https://doi.org/10.1007/s00362-015-0661-7>
- Minds.com, H. (n.d.). Beacon Technology solutions in Retail | Beacon Analytics - Happiest Minds. Retrieved May 12, 2020, from <https://www.happiestminds.com/Insights/Beacons-technology/>
- Müller-Seitz, G., Dautzenberg, K., Creusen, U., & Stromereder, C. (2009). Customer acceptance of RFID technology: Evidence from the German electronic retail sector. *Journal of Retailing and Consumer Services*, 16(1), 31–39. <https://doi.org/https://doi.org/10.1016/j.jretconser.2008.08.002>
- Nies, H. W., Zakaria, Z., Mohamad, Chan, Zaki, N., Sinnott, ... Corchado. (2019). A Review of Computational Methods for Clustering Genes with Similar Biological Functions. *Processes*, 7, 550. <https://doi.org/10.3390/pr7090550>
- Nilsson, R. (n.d.). Bluetooth Low Energy for Wireless Sensors and Actuators. Retrieved May 1, 2020, from <https://www.digikey.com/en/articles/bluetooth-low-energy-for-wireless-sensors-and-actuators>
- Oosterlinck, D., Benoit, D., Baecke, P., & Van de Weghe, N. (2017). Bluetooth tracking of humans in an indoor environment: An application to shopping mall visits. *Applied Geography*, 78, 55–65. <https://doi.org/10.1016/j.apgeog.2016.11.005>
- Pramatari, K., & Theotokis, A. (2011). Consumer Acceptance of RFID-Enabled Services: A Model of Multiple Attitudes, System Characteristics and Individual Traits. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.1329907>
- PWC. (2017). *10 retailer investments for an uncertain future*. Retrieved from <https://www.pwc.com/gx/en/industries/assets/total-retail-2017.pdf>
- retaildive.com. (n.d.). House of Fraser's beacon-enabled mannequins revamp in-store experience. Retrieved from <https://www.retaildive.com/ex/mobilecommercedaily/house-of-frasers-beacon-enabled-mannequins-revamp-in-store-experience>
- Rizos, C., Roberts, G., Barnes, J., & Gambale, N. (2010). Experimental results of Locata: A high accuracy indoor positioning system. In *2010 International Conference on Indoor Positioning and Indoor Navigation* (pp. 1–7). <https://doi.org/10.1109/IPIN.2010.5647717>



- Sano, N., Tsutsui, R., Yada, K., & Suzuki, T. (2016). Clustering of Customer Shopping Paths in Japanese Grocery Stores. *Procedia Computer Science*, 96, 1314–1322. <https://doi.org/10.1016/j.procs.2016.08.176>
- Saura, J. R., Palos-Sánchez, P., & Cerdá Suárez, L. M. (2017). Understanding the Digital Marketing Environment with KPIs and Web Analytics. *Future Internet*, 9(4), 76. <https://doi.org/10.3390/fi9040076>
- Seco, F., Jiménez, A. R., Prieto, C., Roa, J., & Koutsou, K. (2009). A survey of mathematical methods for indoor localization. *WISP 2009 - 6th IEEE International Symposium on Intelligent Signal Processing - Proceedings*, 9–14. <https://doi.org/10.1109/WISP.2009.5286582>
- Senior, A. W., Brown, L., Hampapur, A., Shu, C. F., Zhai, Y., Feris, R. S., ... Carlson, C. (2007). Video analytics for retail. *2007 IEEE Conference on Advanced Video and Signal Based Surveillance, AVSS 2007 Proceedings*, 423–428. <https://doi.org/10.1109/AVSS.2007.4425348>
- Shafique, U., & Qaiser, H. (2014). A Comparative Study of Data Mining Process Models (KDD, CRISP-DM and SEMMA). *International Journal of Innovation and Scientific Research*, 12, 217–222.
- Shearer, C. (2000). The CRISP-DM model: the new blueprint for data mining. *J Data Warehouse*, 5, 13–22.
- SK Telecom. (2015). SK Telecom goes global with indoor location-based services. Retrieved from <https://www.technologyrecord.com/Article/sk-telecom-goes-global-with-indoor-location-based-services-45668>
- Sorensen, H. (2003). The Science of Shopping. *Marketing Research*, 15, 30–35.
- Sorensen, Herb. (2009). The in-store “audience.” *Journal of Advertising Research*, 49(2), 176–179. <https://doi.org/10.2501/S0021849909090242>
- Sorensen, Herb, Bogomolova, S., Anderson, K., Trinh, G., Sharp, A., Kennedy, R., ... Wright, M. (2017). Journal of Retailing and Consumer Services Fundamental patterns of in-store shopper behavior. *Journal of Retailing and Consumer Services*, (February), 1–13. <https://doi.org/10.1016/j.jretconser.2017.02.003>
- Statista.com. (n.d.). • RFID technology market by application worldwide 2014-2025 | Statista. Retrieved April 7, 2020, from <https://www.statista.com/statistics/781314/global-rfid-technology-market-revenue-by-application/#statisticContainer>
- Stavrou, V., Bardaki, C., Papakyriakopoulos, D., & Pramatari, K. (2019). An ensemble filter for indoor positioning in a retail store using bluetooth low energy beacons. *Sensors (Switzerland)*, 19(20). <https://doi.org/10.3390/s19204550>



- Syakur, M., Khotimah, B., Rochman, E., & Dwi Satoto, B. (2018). Integration K-Means Clustering Method and Elbow Method For Identification of The Best Customer Profile Cluster. *IOP Conference Series: Materials Science and Engineering*, 336, 12017. <https://doi.org/10.1088/1757-899X/336/1/012017>
- The Moodie Report. (2016). Schiphol Airport updates app with indoor wayfinding technology. Retrieved from <https://www.moodiedavittreport.com/schiphol-airport-updates-app-with-indoor-wayfinding-technology/>
- Tripathi, S., Bhardwaj, A., & E, P. (2018). Approaches to Clustering in Customer Segmentation. *International Journal of Engineering & Technology*, 7, 802. <https://doi.org/10.14419/ijet.v7i3.12.16505>
- Wang, J., & Loui, M. (2009). Privacy and ethical issues in location-based tracking systems. <https://doi.org/10.1109/ISTAS.2009.5155910>
- Wang, Q., Luo, H., Men, A., Zhao, F., Gao, X., Wei, J., ... Huang, Y. (2018). Light positioning: A high-accuracy visible light indoor positioning system based on attitude identification and propagation model. *International Journal of Distributed Sensor Networks*, 14(2). <https://doi.org/10.1177/1550147718758263>
- Yaeli, A., Bak, P., Feigenblat, G., Nadler, S., Roitman, H., Saadoun, G., ... Sandbank, T. (2014). Understanding customer behavior using indoor location analysis and visualization. *IBM Journal of Research and Development*, 58(5/6), 3:1-3:12. <https://doi.org/10.1147/jrd.2014.2337552>
- Zeimpekis, V., Giaglis, G., & Lekakos, G. (2003). A Taxonomy of Indoor and Outdoor Positioning Techniques for Mobile Location Services. *SIGecom Exchanges*, 3, 19–27. <https://doi.org/10.1145/844351.844355>
- Zhao, J., Wang, L., Li, D., Li, Y., Yang, B., Zhu, B., & Bai, R. (2018). Mining shopping data with passive tags via velocity analysis. *EURASIP Journal on Wireless Communications and Networking*, 2018, 28. <https://doi.org/10.1186/s13638-018-1033-5>
- Γ. ΚΥΡΚΟΣ. (2015). *Επιχειρηματική Ενφύση και Εξόρυξη Δεδομένων*. ΣΥΝΔΕΣΜΟΣ ΕΛΛΗΝΙΚΩΝ ΑΚΑΔΗΜΑΪΚΩΝ ΒΙΒΛΙΟΘΗΚΩΝ. Retrieved from <https://www.kallipos.gr/el/>



Appendix

Appendix A Τεχνικές Λεπτομερείς

Επισημαίνεται ότι για την επεξεργασία των δεδομένων χρησιμοποιήθηκε η γλώσσα προγραμματισμού Python καθώς και βιβλιοθήκες αυτής, ενώ για την αποθήκευση των δεδομένων χρησιμοποιήθηκε η βάση δεδομένων MySQL. Πιο συγκεκριμένα για την υλοποίηση του αλγορίθμου και της ανάλυσης χρησιμοποιήθηκε η διανομή της Python από την Anaconda (<https://www.anaconda.com>) και για περιβάλλον ανάπτυξης (IDE) το Jupyter notebook .

Οι βιβλιοθήκες που χρησιμοποιήθηκαν:

- Για την ανάλυση ήταν οι Pandas, NumPy, Matplotlib, Seaborn, Scikit-Learn, scipy, pyclustering.
- Για τους Γράφους η NetworkX

Επίσης

- Για τους Γράφους έγινε χρήση και του λογισμικού ProM.
- Για τις κατόψεις του καταστήματος χρησιμοποιήθηκε το λογισμικό power bi με το add-on Synoptic Panel.

Appendix B συστήματα εντοπισμού θέσης

Appendix B.1 Τεχνικές εύρεσης θέσης

Αυτή η ενότητα θα παρουσιάσει τεχνικές που χρησιμοποιούνται πιο συχνά στα συστήματα καθορισμού θέσης σε εσωτερικούς χώρους.

- Η τριγωνοποίηση είναι μια μέθοδος όπου η θέση υπολογίζεται από τις γωνίες πρόσπτωσης των ραδιοκυμάτων στο δέκτη. Χρειάζονται τουλάχιστον τρία συστήματα αναφοράς (πομποί ραδιοκυμάτων με γνωστές θέσεις). Εφόσον οι γωνίες μετρηθούν (Angle of Arrival) μπορεί να υπολογιστεί η απόσταση από τα σημεία αναφοράς με τη λύση ενός συστήματος εξισώσεων. Αντίστοιχα δουλεύει και ο τριπλευρισμός όπου αντί να μετριούνται οι γωνίες, μετριούνται οι αποστάσεις (πλευρές του τριγώνου). Αυτές οι τεχνικές, όταν εφαρμόζονται σε εσωτερικούς χώρους, το ηλεκτρομαγνητικό σήμα παίρνει πολλές, διαφορετικές διαδρομές γιατί ανακλάται από μεταλλικές επιφάνειες ή χάνει ισχύ όταν περνά μέσα από συμπαγή υλικά για να φτάσει στο δέκτη. Αυτό οδηγεί σε εσφαλμένες εκτιμήσεις τοποθεσίας (Stavrou et al., 2019).
- Η Fingerprinting είναι μια μέθοδος που χρησιμοποιείται για την εκτίμηση της θέσης. Η βασική ιδέα αυτής της προσέγγισης είναι να χρησιμοποιηθεί η ένταση του σήματος RSSI για να εντοπιστεί η θέση του καταναλωτή, καθώς το κατάστημα χωρίζεται σε περιοχές, και σε κάθε περιοχή τοποθετούνται beacons συσκευές. Πλεονέκτημα της μεθόδου είναι ότι δεν χρειάζεται να γνωρίζουμε τις θέσεις των beacons στο κατάστημα. Η μέθοδος αποτελείται από δύο φάσεις :
 - Την Offline, κατά την οποία δημιουργείται ένας χάρτης με τις εντάσεις του σήματος από τα beacons εντός του καταστήματος, Το αποτέλεσμα είναι η κάθε περιοχή μέσα στο κατάστημα να καθορίζεται από τις διαφορετικές μετρήσεις του RSSI, οι οποίες αποθηκεύονται σε μια βάση δεδομένων και είναι μοναδικές για κάθε περιοχή
 - Την online, κατά την οποία ο πελάτης κινείται μέσα στο κατάστημα, η συσκευή του συλλέγει μετρήσεις του RSSI, τις συγκρίνει με τις μετρήσεις που είναι αποθηκευμένες στην βάση δεδομένων και είναι ταυτισμένες με μια περιοχή.

Appendix B.2 Μετρικές σύγκρισης των συστημάτων εντοπισμού

Οι ακόλουθες μετρικές όπως τις ορίσαν οι Liu et al. χρησιμοποιούνται για τον υπολογισμό της απόδοσης ενός εσωτερικού συστήματος εντοπισμού.

- **Ακρίβεια:** Η ακρίβεια(accuracy ή localization error) είναι η πιο σημαντική μέτρηση απόδοσης για ένα σύστημα εντοπισμού θέσης. Συνήθως το μέσο σφάλμα απόστασης χρησιμοποιείται ως μέτρηση απόδοσης, δηλαδή η ευκλείδεια απόσταση μεταξύ της υπολογιζόμενης θέσης και της πραγματικής θέσης. Όσο υψηλότερη είναι η ακρίβεια τόσο καλύτερο θεωρείται το σύστημα, ωστόσο υπάρχει πάντα μια αντιστάθμιση μεταξύ ακρίβειας και τα αλλά χαρακτηριστικά του συστήματος. Επίσης η επιθυμητή ακρίβεια σχετίζεται πάντα το σκοπό χρήση του συστήματος ακριβή εντοπισμό σε επίπεδο δωμάτιου και διάδρομου απαιτούν διαφορετική ακρίβεια ακόμα και διαφορετική υλοποίηση .
- **Ορθότητα:** Ενώ η ακρίβεια περιγράφει την τιμή του μέσου σφάλματος απόστασης, η Ορθότητα για να περιγράψει πόσο σταθερά λειτουργεί το σύστημα, δείχνει τη διακύμανση των επιδόσεών του σε πολλές δοκιμές. Μπορεί να περιγραφεί ως η κατανομή του σφάλματος απόστασης μεταξύ της εκτιμώμενης θέσης και της πραγματικής συνήθως χρησιμοποιεί η αθροιστική συνάρτηση κατανομής CDF για την απεικόνιση.
- **Πολυπλοκότητα:** Η πολυπλοκότητα ενός συστήματος εντοπισμού αποδίδεται στις υλικοτεχνικές υποδομές που χρειάζονται για την υλοποίηση του καθώς και την Εφαρμογή άλλα στοιχεία του συστήματος.
- **Κόστος:** Το κόστος ενός συστήματος εντοπισμού εξαρτάται από πολλούς παράγοντες. Σημαντικότεροι παράγοντες μεταξύ άλλων είναι Χρηματικό κόστος, ο χρόνο (συντελεστής χρόνου σχετίζεται με την εγκατάσταση και συντήρηση του συστήματος στην περίπτωση της τεχνικής fingerprinting χρειάζεται να δημιουργηθεί ο τοπογραφικός χάρτης ο οποίος χρειάζεται και προσαρμογή που και που αν θέλουμε να μιλήσουμε για ένα RTLS να εντοπίζει σε πραγματικό χρόνο τον χρήστη και να αλληλεπιδρά μαζί του), ο φυσικός χώρος που καταλαμβάνει ένα beacon ή rfid , το βάρος (οι κινητές μονάδες ενδέχεται να έχουν περιορισμούς βάρους και διαστάσεων για να μπορούν να τοποθετηθούν και να μην επηρεάζει/παρεμποδίζει τον χρήστη) και η ενέργεια.
- **Ανθεκτικότητα:** Ένα σύστημα εντοπισμού με υψηλό βαθμό ευρωστίας θα μπορούσε να λειτουργήσει σωστά ακόμη και σε ένα σενάριο όπου ορισμένοι πομποί (beacons/AC) δεν είναι διαθέσιμοι ή δεν έχουν ξαναδεί ποτέ μερικές από τις τιμές τους όπως RSS ή μιας γωνίας.
- **Πρακτικότητα:** Αυτή η μέτρηση σημαίνει πόσο πρακτική είναι η επιλογή των προσεγγίσεων εντοπισμού να αναπτύξει το σύστημα εντοπισμού για ένα συγκεκριμένο εσωτερικό περιβάλλον. Αυτό συνεπάγεται πόσο πρακτική είναι η ανάπτυξη του επιλεγμένου υλικού, η



εφαρμογή του αλγορίθμου εντοπισμού και η συνολική λειτουργικότητα των συστημάτων σε δυναμικά εσωτερικά περιβάλλοντα.

Συμπεράσματα

Από την έρευνα προκύπτει ότι υπάρχουν διάφορες αποτελεσματικές μέθοδοι καθορισμού θέσης σε εσωτερικούς χώρους οι οποίες έχουν τα πλεονεκτήματα και τα μειονεκτήματα τους. Ο κατωτέρω πίνακας παρουσιάζει μια σύντομη αξιολόγηση για κάθε μέθοδο.

•	• GPS	• Wi-fi	• RFID	• BLE Beacons
• Ορθότητα	• Υψηλή(εξωτερικά) • Πολύ χαμηλή (εσωτερικά)	• Υψηλή	• Μεσαία	• Υψηλή
• Ακρίβεια	• Υψηλή(εξωτερικά) • Πολύ χαμηλή (εσωτερικά)	• Μεσαία (εξαρτάται από κάλυψη και αριθμό των AC)	• Μεσαία	• Υψηλή
• Πολυπλοκότητα	• Χαμηλή	• Υψηλή	• Υψηλή	• Χαμηλή
• Υποστήριξη από κινητό τηλέφωνο	• Ναι	• Ναι	• Όχι	• Ναι
• Απαιτείται εξοπλισμό	• Ναι	• Όχι •	• Ναι	• Ναι
• Κόστος	• Χαμηλό/Υψηλό	• Χαμηλό	• Υψηλό	• Χαμηλό

Είναι ξεκάθαρο ότι είτε το Wi-fi είτε τα beacons είναι οι καλύτερες επιλογές για ακριβή και αξιόπιστη εύρεση θέσης σε εσωτερικούς χώρους (LBS). Το GPS είναι ακατάλληλο για εξωτερικούς χώρους και τα RFID είναι ακριβό, πολύπλοκο και δεν παρέχει την ίδια ακρίβεια με το Wi-fi ή το BLE Beacons. Το Wi-fi έχει καλή ακρίβεια με χαμηλό κόστος αλλά είναι συνήθως πιο πολύπλοκο να εφαρμοστεί. Τα BLE Beacons είναι πιο εύκολα να εφαρμοστούν, παρέχουν άριστη ορθότητα και ακρίβεια και είναι πολύ καλύτερη επιλογή από άποψη κόστους. Δεδομένου της επικράτησης των συσκευών Bluetooth όλες οι κινητές συσκευές διαθέτουν Bluetooth ακόμα και οι περισσότερες έξυπνες συσκευές οπός αποκαλούνται επικοινωνούν είτε με WIFI ή με Bluetooth.

Appendix C

Πίνακα Πιθανοτήτων Μετάβασης οπός αυτός χρησιμοποιήθηκε για να παραχθεί η εικόνα των γράφων στο Κεφάλαιο 4 Κάθε διάνυσμα-γραμμή $P^{(o)}$ αποτελεί κατανομή των αρχικών καταστάσεων της αντίστοιχης κατηγορίας $P_j^{(o)}$. Επειδή είναι τετραγωνικός πίνακας P με μη αρνητικά στοιχεία $p_{ij} \geq 0$ με $\sum_j p_{ij} = 1$ ονομάζεται Στοχαστικός και αποτελεί Πίνακα Πιθανοτήτων Μετάβασης μιας Μαρκοβιανής αλυσίδας.

Sum of probabilities	Column L													
Row Labels	accessories	cardigan	cash registers	cloak	cloak+1 accessory	coat	dress	dress+1 accessory	dungaree	gilet	jacket	jacket+1 accessory	jacket+dress	
accessories	0,089791	0,241325927				0,091613	0,0978053	0,093616246						
cardigan	0,6018018												0,330673594	
cloak													0,256284011	
cloak+1 accessory														
coat													0,3781175	
dress	0,62188254													
dress+1 accessory	0,54764065													
dungaree	0,54878049													
gilet							0,251658							
jacket					0,324849297									
jacket+1 accessory														
jacket+dress			0,489025											
jacket+skirt			0,47387											
jacket+trousers				0,460680128					0,5393199					
longtrousers	0,253399												0,186402266	
mid-lengthjacket	0,51759062													
mid-lengthjacket+skirt	0,60460913													
mid-lengthjacket+trousers	0,51258346								0,4874165					
overcoat+raincoat		0,237765	0,249608939	0,259									0,253407821	
packofshirt										0,470780739				
packoftops	0,52550364													
shirt										0,498489426				
shorttrousers				0,156					0,1604069				0,156298905	
skirt			0,463627049						0,536373					
top								0,198489373	0,236079396	0,1949763				
Grand Total	4,48039234	0,343191	0,241325927	1,20066	1,498765412	0,415	0,343271	0,296294673	0,329695642	2,29661	1,225554176	0,409706726	0,517075861	
	jacket+skirt	jacket+trousers	longtrousers	mid-lengthjacket	mid-lengthjacket+skirt	mid-lengthjacket+trousers	overcoat+raincoat	packofshirt	packoftops	shirt	shorttrousers	skirt	top	Grand Total
				0,099353429	0,092796649	0,09680357		0,096894636						1
				0,398198198										1
i 0,326034515	0,021154203				0,322137688									1
	0,230370562				0,235294118				0,278051					1
					0,451662404				0,548337596					1
										0,452359				1
										0,45122				1
	0,149464377				0,150314572				0,145553477	0,149464	0,153545			1
							0,357334226	0,3178						1
							0,490429384			0,509570616				1
														1
				0,510974822										1
				0,52613017										1
i 0,192917847														1
	0,482409382									0,187535411				1
										0,395390872				1
														1
														1
				0,474496357										1
														1
				0,140649452	0,15356025	0,233568075								1
									0,188301423	0,1822				1
l 0,518952362	0,400989142	2,532858381	0,252913679	0,326364724	0,247118142	1,499523594	0,545635649	0,276639679	0,5	1,786387972	0,427516	2,087854		25

Πίνακας 9: Στοχαστικός πίνακας των ποσοστών μετακίνησης/μετάβασης από μια (αρχική κατάσταση) κατηγορία (γράμμη) σε άλλη (τελική κατάσταση) κατηγορία (στήλη)

