ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ΣΧΟΛΗ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ

ΤΜΗΜΑ ΔΙΟΙΚΗΤΙΚΗΣ ΕΠΙΣΤΗΜΗΣ ΚΑΙ ΤΕΧΝΟΛΟΓΙΑΣ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

του

ΑΝΔΡΕΑ ΠΛΕΣΣΑ

Data mining techniques σε Online Κατάστημα στην Ελλάδα

Επιβλέπων : Γιώργος Λεκάκος - Καθηγητής/Ζάρας Ανδρέας

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

Αθήνα, Μάρτιος 2022



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)

.....

ΠΛΕΣΣΑΣ ΑΝΔΡΕΑΣ

Φοιτητής MSc στη Διοικητική Επιστήμη και Τεχνολογία



Ευχαριστίες

Θα ήθελα να ευχαριστήσω την γυναίκα μου Ερμιόνη που με στήριξε και με στηρίζει πάντα. Επίσης, την μητέρα μου Αθηνά, που πάντα με ωθούσε να επιτύχω όσα περισσότερα μπορούσα. Επιπλέον, να ευχαριστήσω την Στέλλα Κουλουλία, η οποία με το Viva της μου έδωσε έμπνευση και τον φίλο μου Δημήτρη, που χωρίς την δική του καθοριστική συμβολή δεν θα ήταν εφικτή η εκπόνηση της διπλωματικής. Τέλος να ευχαριστήσω τον κ. Ζάρα για την πολύτιμη βοήθειά του και την καθοδήγησή του.

Αφιερωμένη στην θεία μου Σπυρούλα
που έφυγε πριν 10 μέρες από κοντά μας.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Περίληψη

Με την εξέλιξη της τεχνολογίας δημιουργούνται νέες δυνατότητες για το Επιχειρείν. Παράλληλα, η αυξανόμενη συλλογή, συγκέντρωση και αποθήκευση δεδομένων από τους οργανισμούς σε σχέση με την εξέλιξη των τεχνικών επεξεργασίας τους, ανοίγουν νέους δρόμους στη δημιουργία δυναμικών θεωριών για την προσέλκυση των νέων πελατών και διατήρησης των υφιστάμενων. Τα ηλεκτρονικά καταστήματα ανθίζουν στην Ελλάδα τα τελευταία χρόνια, έχοντας δημιουργήσει έναν υγιή ανταγωνισμό. Ο ανταγωνισμός αυτός αντικατοπτρίζεται στις ενέργειες κάθε οργανισμού να προσελκύσει μεγαλύτερο μέρος πελατών και κατά συνέπεια, να κατακτήσει μεγαλύτερο μερίδιο αγοράς. Τα Data Mining Techniques έρχονται να εφαρμόσουν νέες λειτουργικότητες με δυνατότητες, οι οποίες μπορούν να απογειώσουν τα έσοδα μιας επιχείρησης. Πρώτα από όλα, ο Οργανισμός πρέπει να γνωρίζει το πελατολόγιο του και η απάντηση σε αυτό είναι το Customer Segmentation., ώστε να μπορέσει να εφαρμόσει εξατομικευμένες στρατηγικές Marketing. Αφού διαχωριστούν σε clusters οι πελάτες, θα πρέπει να γνωρίζει ποιες είναι οι προτιμήσεις τους. Οι τεχνικές RFM Analysis και το Market Basket Analysis αποτελούν δημοφιλείς τεχνικές που έρχονται να απαντήσουν σε καίρια τέτοια ερωτήματα και ακόμα περισσότερο να δημιουργήσουν νέα.

Λέξεις Κλειδιά: <electronic commerce, consumer behaviour, customer services, marketing, marketing data processing, pattern clustering, purchasing, retail data processin, market basket analysis, RFM analysis, k-means clustering, >>

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Η σελίδα αυτή είναι σκόπιμα λευκή



Πίνακας περιεχομένων

1 Εισαγωγή 1

1.1 Big Data στο χώρο των Marketing Data Analytics 1

1.2 E-shops 4

1.3 Αντικείμενο διπλωματικής 7

1.3.1 Συνεισφορά 8

1.4 Οργάνωση κειμένου 8

2 Σχετικές εργασίες 9

2.1 Customer Segmentation - Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining 9

2.2 Association Rules - Finding Ideal Menu Items Assortments: An Empirical Application of Market Basket Analysis 9

3 Θεωρητικό υπόβαθρο 10

3.1 Digital Marketing Strategies 11

3.1.1 Στρατηγικές Digital Marketing 11

3.2 Data exploration 12

3.2.1 Δειγματοληψία 14

3.2.2 Κατηγορίες δεδομένων 15

3.2.3 Data Cleansing 15

3.3 Data Mining 16

3.4 RFM Model (Recency, Frequency, and Monetary) 18

3.4.1 Πλεονεκτήματα και μειονεκτήματα χρήσης RFM 20

3.4.2 RFM Μεθοδολογία 21

3.4.3 Segmentation 22

3.5 Market Basket Analysis 23

3.5.1 Πλεονεκτήματα χρήσης Market Basket Analysis 24

3.5.2 Εφαρμογές 25

3.5.3 Market Basket Analysis ορισμοί 26

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

3.5.4 Consumer Behaviour *28*

3.5.5 General Data Protection Regulation (GDPR) *29*

4 Πώς οι τεχνικές data analytics συμβάλουν στην αύξηση των εσόδων των ηλεκτρονικών καταστημάτων **32**

5 Η Επιχείρηση **34**

5.1 Business Background 34

5.2 Παρουσίαση Dataset 35

5.2.1 Data Set Customers *35*

5.2.2 Data Set Orders Items *35*

5.2.3 Data Set Orders payments analysis *36*

5.2.4 Data Set Payment method *37*

5.3 Ανάλυση Πελατολογίου 37

5.4 Πωλήσεις 41

6 Εφαρμογή RFM Analysis - Market Basket Analysis και Αξιολόγηση αποτελεσμάτων **43**

6.1 RFM Model 43

6.1.1 Παράμετροι αξιολόγησης *43*

6.1.2 Επεξήγηση των Clusters *46*

6.1.3 Segment Profiling *47*

6.1.1 Αποτελέσματα *47*

6.1.2 Σύνοψη συμπερασμάτων αξιολόγησης *48*

6.2 Market Basket Analysis 49

6.2.1 Μεθοδολογία MBA *51*

6.2.2 Αποτελέσματα *53*

6.2.3 Σύνοψη συμπερασμάτων αξιολόγησης *56*

7 Τεχνικές λεπτομέρειες **58**

7.1 Πλατφόρμες και προγραμματιστικά εργαλεία 58

7.2 Λεπτομέρειες υλοποίησης 58

7.2.1 Μοντέλο RFM *58*

7.2.2 Market Basket Analysis *60*

8 Επίλογος **63**

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

9 Bibliography 65

Appendices 67

Λίστα πινάκων 67

Λίστα Charts 67

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

1

Εισαγωγή

1.1 Big Data στο χώρο των Marketing Data Analytics

Μπορούμε να ισχυριστούμε ότι τα δεδομένα γεννήθηκαν πριν την εμφάνιση του ανθρώπου. Τη στιγμή που η αμοιβάδα, από τους πρώτους οργανισμούς που εμφανίστηκαν στη γη, διαιρείται προκειμένου να πολλαπλασιαστεί, τότε παράγεται η πληροφορία της διαδικασίας της αναπαραγωγής. Με αυτό τον τρόπο, όλος ο κόσμος πλημμυρίζεται από δεδομένα. Κάθε κίνηση, κάθε στιγμή αποτελεί μια πληροφορία και μεταφράζεται σε δεδομένα. Σήμερα τεράστιοι όγκοι δεδομένων παράγονται μέσω πηγών όπως Personal Computers, Smartphones, Social Networks κτλ (Klein, 2017).

Κάθε μήνα λαμβάνουμε ηλεκτρονική αλληλογραφία από την Τράπεζα που συνεργαζόμαστε με επισυναπτόμενο αρχείο που περιέχει τις αναλυτικές κινήσεις του τραπεζικού μας λογαριασμού. Παράλληλα, στο μηνιαίο αναλυτικό λογαριασμό του τηλεφώνου μας περιέχονται οι εξερχόμενες κλήσεις που έχουμε πραγματοποιήσει με αναγραφόμενο τον αριθμό που καλέσαμε, τη διάρκεια κλήσης και το κόστος. Στο τέλος της ημέρας, επισκεπτόμαστε το super market στο οποίο διατηρούμε Loyalty card και οι αγορές που πραγματοποιούμε καταγράφονται στον λογαριασμό όπου θα μεταφραστούν σε bonus πόντους με δυνατότητα εξαργύρωσης.

Όπως ευκολά αντιλαμβανόμαστε τα δεδομένα μας περιτριγυρίζουν και είναι μια πραγματικότητα που ήρθε για να μείνει. Αυτός ο κόσμος αλλάζει εκπληκτικά γρήγορα και τα δεδομένα μπορούν να

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

απεικονίσουν το πόσο. Για πολλά χρόνια η επεξεργασία των δεδομένων περιοριζόταν σε εκείνα που δημοσίευαν επίσημα οι εταιρίες Στατιστικής και ερευνών για να αξιολογήσουν την κατάσταση μιας οικονομίας. Η συλλογή των δεδομένων συνήθως ήταν χρονοβόρα και δαπανηρή και συχνά απαιτούσε σημαντική προσπάθεια. Ωστόσο, με την εξέλιξη της τεχνολογίας έχει παρατηρηθεί τα τελευταία χρόνια μια σημαντική αύξηση του όγκου των δεδομένων που παράγονται, ενώ είναι ευκολότερη η εξαγωγή και η ανάλυσή τους. Σε αυτό συνέβαλε η δημιουργία εργαλείων που δίνουν την δυνατότητα να πραγματοποιηθεί επεξεργασία μεγάλου όγκου σε πραγματικό χρόνο. Τεχνολογίες όπως τα Smart Phones, το Blockchain και το Cloud Computing αποτελούν παράγοντες μετάδοσης της πληροφορίας. Κατά συνέπεια, η συλλογή πληροφοριών από χρήστες στο Διαδίκτυο παράγει συμπεράσματα για τις προτιμήσεις αγοράς των χρηστών σε ένδυση, διασκέδαση και υπηρεσίες. Οι επιχειρήσεις έχουν την δυνατότητα να παρακολουθήσουν την παραγωγή, την επεξεργασία και την πώληση των προϊόντων/υπηρεσιών τους.

Ο νέος όρος που χρησιμοποιείται για όλες τις διάχυτες ανά τον κόσμο πληροφορίες είναι Big Data. Ποια είναι όμως τα Big Data; Είναι όλα τα μη δομημένα, ημι-δομημένα ή δομημένα Data που συλλέγονται από οργανισμούς. Για την εξόρυξη τους και στη συνέχεια για την επεξεργασία τους χρησιμοποιούνται ειδικές τεχνικές και εργαλεία όπως machine learning, predictive models και άλλα εργαλεία.

Ειδικότερα, τα δεδομένα αποτελούν αναπόσπαστο κομμάτι της λειτουργίας των επιχειρήσεων. Σε παγκόσμιο επίπεδο όλες οι εταιρίες προσπαθούν καθημερινά να διαχειριστούν τεράστιο όγκο δεδομένων και αρκετές φορές αγνοούν το γεγονός ότι το πράττουν. Η επεξεργασία τους με σκοπό της παραγωγή αποτελεσμάτων δίνει ανταγωνιστικό πλεονέκτημα και ευκαιρίες δημιουργίας στρατηγικών διαχείρισης για κάθε οργανισμό. Μπορούν να χρησιμοποιηθούν για τη βελτίωση των λειτουργιών, την παροχή ποιοτικότερης εξυπηρέτησης πελατών και για την διεξαγωγή Marketing Campaigns. Δημιουργεί νέες δυνατότητες για παροχή ποιοτικότερων υπηρεσιών και νέα προϊόντα προς τους πελάτες με τελικό σκοπό την παραγωγή κέρδους για τις επιχειρήσεις. Παράλληλα, μας βοηθούν να έχουμε πρόσβαση σε συγκεκριμένες αναλυτικές προσεγγίσεις ώστε να είναι εφικτή η λήψη αποφάσεων.

Τα Big Data ορίζονται από τα χαρακτηριστικά που τα διακρίνουν. Το 2001 ο (Douglas) όρισε τα χαρακτηριστικά γνωστά ως και 3 Vs, το Volume, το Velocity και το Variety.

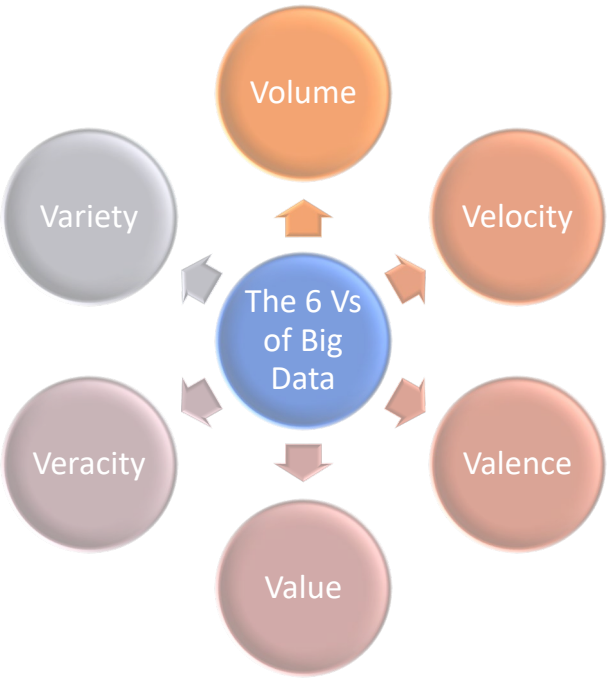
Τα Big Data ορίζονται από τα χαρακτηριστικά που τα διακρίνουν. Το 2001 ο (Douglas) όρισε τα χαρακτηριστικά γνωστά ως και 3 Vs, το Volume, το Velocity και το Variety.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

- **Volume** – (Όγκος) ο όγκος δεδομένων παίζει σημαντικό ρόλο και εφόσον είναι μεγάλος τότε θεωρούνται Big Data. Η εξέταση του όγκου αποτελεί μέρος της επεξεργασίας.
- **Velocity** – (Ταχύτητα) αφορά στην ταχύτητα με την οποία τα δεδομένα εισέρχονται και εξέρχονται από ένα σύστημα, αλλά και την ταχύτητα με την οποία αυτά τυγχάνουν επεξεργασίας και ανάλυσης από αυτό.
- **Variety** – (Ποικιλομορφία) αφορά στην ποικιλία των τύπων (Structured, Semi-Structured, Unstructured) και της φύσης των δεδομένων (Documents, Databases, Images, Videos, Logs, κλπ.), αλλά και το πλήθος των διαφόρων πηγών τους.

Αργότερα, προστέθηκαν επιπλέον 3 Vs στην περιγραφή των Big Data:

- **Veracity** – Αφορά στην ακρίβεια των δεδομένων, την αξιοπιστία της πηγής δεδομένων και το πλαίσιο της ανάλυσής τους
- **Valence** – Μεταξύ των δεδομένων υπάρχει συνδεσιμότητα και αυτό συμβαίνει όταν σχετίζονται μεταξύ τους
- **Value** - (Αξία) αφορά στα δυνητικά έσοδα, οικονομικά και κοινωνικά, που μπορεί να προκύψουν από την εκμετάλλευσή τους. Τα δεδομένα που δεν επεξεργάζονται δεν έχουν καμία αξία. Όταν όμως τεθούν προς επεξεργασία, τότε αποκτούν αξία για κάθε οργανισμό.



Εικόνα 1 - 6Vs of Big Data

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

1.2 E-shops

Η ανάπτυξη της τεχνολογίας, η εξέλιξη των αυτοματισμών και η ψηφιοποίηση των υπηρεσιών έχουν μεταβάλλει τη ζωή των πολιτών στα περισσότερα κράτη του κόσμου. Κατά τη μετάβαση στη νέα πραγματικότητα παρατηρείται η σταδιακή εξασθένιση επαγγελμάτων, αλλά και η επικράτηση νέων που σχετίζονται περισσότερο με την τεχνολογία. Καθώς η τεχνολογία συνεχίζει να εξελίσσεται, αποτελεί μια μεταμορφωτική δύναμη, η οποία αλλάζει τον τρόπο που τα άτομα ζουν, αλληλεπιδρούν και εργάζονται. Σήμερα το ηλεκτρονικό εμπόριο έχει γίνει αναπόσπαστο κομμάτι της καθημερινότητας. Η προσβασιμότητα σε πλατφόρμες ηλεκτρονικού εμπορίου δεν είναι προνόμιο, αλλά μάλλον αναγκαιότητα για τους περισσότερους ανθρώπους, ιδιαίτερα στις αστικές περιοχές. (Kumar V., 2014). Το ηλεκτρονικό εμπόριο καλύπτει το μεγαλύτερο μέρος των κλάδων, όπως αγορές ηλεκτρονικών και οικιακών συσκευών, προϊόντα διατροφής, χρηματοοικονομικές συναλλαγές, ακόμα και μετοχές. Βασικός μέτοχος της ψηφιοποίησης είναι η αλματώδης ανάπτυξη των τεχνολογικών εφαρμογών που μας επιτρέπουν να έχουμε πρόσβαση στο Διαδίκτυο.

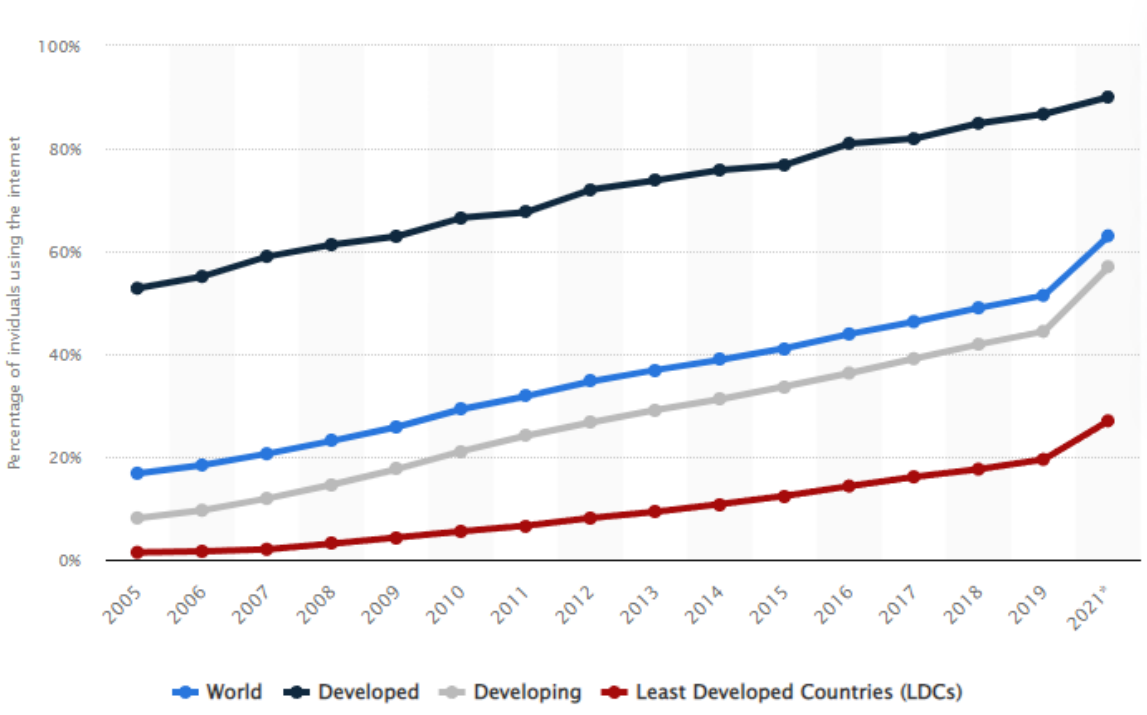


Chart 1 - Χρήση Internet 2005 -2021 (Johnson, January 2022)

Στο παραπάνω γράφημα απεικονίζεται η πορεία της χρήση του Διαδικτύου για την περίοδο 2005-2019 και η πρόβλεψη για το 2021 από την οποία προκύπτει το συμπέρασμα ότι περισσότερο από το 60% παγκοσμίως θα χρησιμοποιεί το Internet. Ειδικότερα, φαίνεται ότι το 90% των πολιτών σε αναπτυγμένες χώρες έχει πρόσβαση και χρησιμοποιεί το διαδίκτυο για διαφορετικές ανάγκες σε

σύγκριση με το ποσοστό των αναπτυσσόμενων χωρών το οποίο είναι μικρότερο του 60%. Στις λιγότερο ανεπτυγμένες χώρες, το 2021 εκτιμάται ότι μόλις το 27% θα έχει πρόσβαση στο Διαδίκτυο. Διαπιστώνεται, επίσης, ότι σε βάθος δεκαετίας η χρήση του διαδικτύου έχει αυξηθεί περίπου στο 100%, δημιουργώντας με αυτό τον τρόπο διπλάσιους χρήστες.

Η εμφάνιση των πρώτων ηλεκτρονικών καταστημάτων χρονολογείται στην έναρξη της δεκαετίας του 90. Αρχικά, οι επιχειρήσεις μέσω του παγκόσμιου ιστού παρουσίαζαν τα προϊόντα ή τις υπηρεσίες τους και μπορούμε να θεωρήσουμε ότι ήταν τα πρώτα βήματα που οδήγησαν στη δημιουργία των σημερινών e-shops. Στη συνέχεια, ξεκίνησαν να δέχονται παραγγελίες μέσω αλληλογραφίας και επέστρεφαν στους ενδιαφερόμενους πληροφορίες και οδηγίες χρήσης των προϊόντων ή των υπηρεσιών. Ακολούθησε η εισαγωγή της ψηφιακής τεχνολογίας για τη μεταφορά των συγκεκριμένων εμπορευμάτων, προσφέροντας τη δυνατότητα στον καταναλωτή να γνωρίζει τον χρόνο παράδοσης της αγοράς του. Πλέον κάθε καταναλωτής απολαμβάνει μια εμπειρία σε κάθε αγορά του αλληλοεπιδρώντας με τον πωλητή, ενώ λαμβάνει ακόμα και προτάσεις αγοράς με βάση την προηγούμενη ή την επικείμενη αγορά του.

Στις αρχές του 2020 ήρθαμε αντιμέτωποι με τη μεγαλύτερη δοκιμασία της σύγχρονης εποχής: την εξάπλωση μιας πανδημίας, της ασθένειας γνωστή με την ονομασία Covid-19. Καθώς οι επιπτώσεις της ασθένειας ήταν τεράστιες και η εξάπλωσή της ραγδαία σε όλες της χώρες του κόσμου, πολλά κράτη έλαβαν άμεσα την απόφαση απαγόρευσης κυκλοφορίας με στόχο το περιορισμό έκθεσης των πολιτών στον ιό. Σε πολύ σύντομο χρονικό διάστημα η καθημερινότητα εκατομμυρίων ανθρώπων άλλαξε τελείως και απλές, καθημερινές συνήθειες, όπως η επίσκεψη σε ένα super market ή ένα κατάστημα λιανικού εμπορίου, έγιναν πολύπλοκες. Καθώς η πραγματικότητα μεταβλήθηκε αστραπιαία, πολλές επιχειρήσεις αναγκάστηκαν να παραμείνουν κλειστές για αρκετό διάστημα, ανάλογα με τα μέτρα που έλαβε κάθε χώρα. Όπως όμως έχει αποδειχθεί στον ρου της ιστορίας, όταν οι συνθήκες της αγοράς μεταβάλλονται, νέες ευκαιρίες δημιουργούνται. Για παράδειγμα, η παραγωγή των ιατρικών μασκών που χρησιμοποιούνται για την προστασία από τον κορωνοϊό αυξήθηκε τόσο πολύ, που πλέον η βιομηχανική δύναμη δεν επαρκούσε να καλύψει τις ανάγκες ζήτησης. Ακόμα περισσότερο, ο κλάδος των μεταφορών σημείωσε ταχύτατα μεγάλη άνοδο στη ζήτηση και την προσφορά των υπηρεσιών, αφού η αδυναμία φυσικής παρουσίας του καταναλωτή σε φυσικό κατάστημα είχε ως αποτέλεσμα την αύξηση τηλεφωνικών και ηλεκτρονικών παραγγελιών. Ιδιαίτερα, στις περιόδους των λεγόμενων lockdown, οι καταναλωτές έχοντας περισσότερο χρόνο στην διάθεσή τους διότι έμεναν το μεγαλύτερο μέρος της ημέρας στο σπίτι τους, αναζητούσαν στο Διαδίκτυο προϊόντα και υπηρεσίες για τις καθημερινές και μη ανάγκες τους. Συνακόλουθα, τα καταστήματα αυξήθηκαν κατακόρυφα, καθώς η δημιουργία e-shops ήταν επιτακτική λύση για τις επιχειρήσεις online. Πλέον κάθε φυσικό κατάστημα προκειμένου να επιβιώσει στις νέες συνθήκες, λειτουργεί παράλληλα ένα ηλεκτρονικό κατάστημα.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Για να γίνει ευκολότερα αντιληπτή η εξέλιξη της πορείας των ηλεκτρονικών καταστημάτων, αντλήσαμε στοιχεία από την Eurostat, τη στατιστική υπηρεσία της Ευρωπαϊκής Επιτροπής για να αναδειχθεί η εξέλιξη που έχει σημειωθεί την τελευταία δεκαετία. Αναλυτικότερα στο παρακάτω γράφημα απεικονίζει τα μεγέθη σε group ηλικιών που πραγματοποιούν αγορές μέσω Διαδικτύου. Παρατηρείται αλματώδης αύξηση 60% στους χρήστες ηλικιών 16-24, ενώ για το group 25-54 η αύξηση είναι πιο ομαλή σημειώνοντας μέγιστο όριο τα 81%. Τέλος το group ηλικιών έχει την μικρότερη δυνατή αύξηση λίγο λιγότερη από 60%

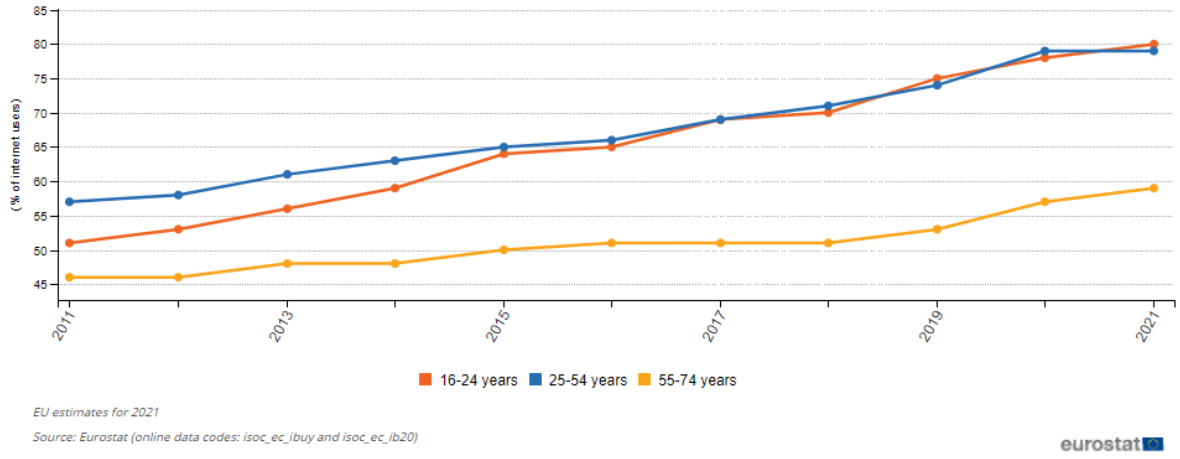


Chart 2 - Internet users who bought or ordered goods or services for private use in the previous 12 months by age group, EU, 2011-2021

Στο επόμενο Chart περιγράφεται η αυξητική τάση των πωλήσεων σε παγκόσμιο επίπεδο με πρόβλεψη μέχρι το έτος το 2024 να ξεπερνούν τα 6 δις σε δολάρια.

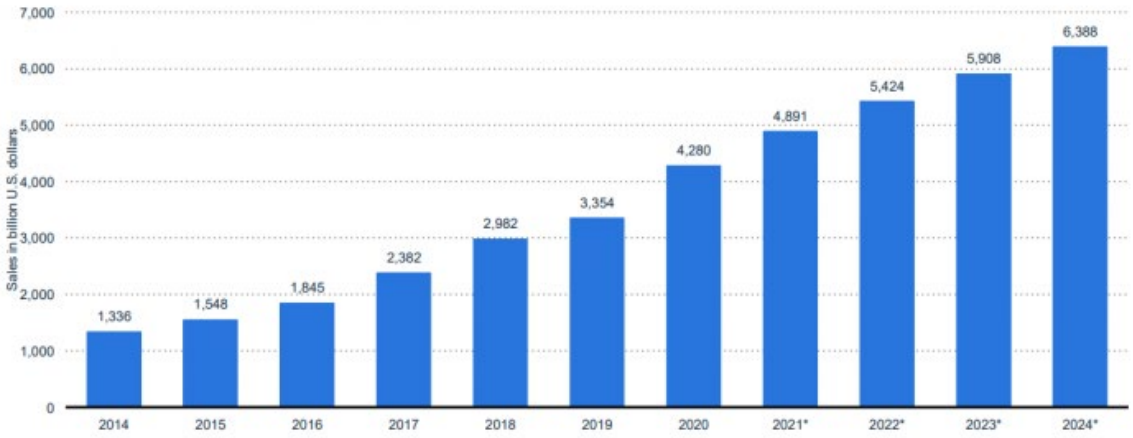


Chart 3 - Global retail ecommerce sales worldwide from 2014 to 2024 (in billion USD)

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Τα ηλεκτρονικά καταστήματα είναι σχεδόν απίθανο να αντικαταστήσουν τα φυσικά καταστήμα. Το μόνο σίγουρο είναι ότι κατέχουν μεγάλο μερίδιο της αγοράς, το οποίο θα αυξηθεί δρματικά τα επόμενα χρόνια. Όμως μαζί με την αυξηση των καταστημάτων, αυξάνεται και η παραγωγή δεδομένων από τις επιχειρηματικές ενέργειες. Για να μπορέσουν οι επιχειρήσεις να παραμείνουν ανταγωνιστικές πρέπει να εκμεταλλευτούν αυτά τα δεδομένα. Η επεξεργασία και η ανάλυση τους θα προσφέρουν πολύ χρήσιμα συμπεράσματα για τις επιχειρήσεις και θα μπορέσουν να λάβουν καλύτερες αποφάσεις για το μέλλον. Ερωτήματα όπως ποια προιόντα αγοράζουν οι πελάτες, ποιοι πελάτες αξίζουν πιο πολύ για την επιχείρηση μπορούν να απαντηθούν με την επεξεργασία των Big Data εφαρμόζοντας τεχνικές Data Mining.

1.3 Αντικείμενο διπλωματικής

Για τις ανάγκες της διπλωματικής αντικείμενο μελέτης είναι ένα ηλεκτρονικό κατάστημα στη Ελλάδα. Η εύρεση επιχείρησης διαθέσιμης να προσφέρει δεδομένα προς επεξεργασία για την παραγωγή αποτελεσμάτων ήταν αρκετά δύσκολη. Τα τελευταία χρόνια με την εφαρμογή του Γενικού Κανονισμού Προστασίας Δεδομένων της Ε.Ε. ή όπως είναι γνωστός ως GDPR (General Data Protection Regulation), οι επιχειρήσεις είναι υποχρεωμένες να τον τηρούν και είναι ιδιαιτέρως επιφυλακτικές στην παροχή δεδομένων, ακόμα και για εκπαιδευτικούς σκοπούς. Η επιχείρηση, η οποία συμφώνησε να συνδράμει με τα πραγματικά της δεδομένα στην εκπόνηση της διπλωματικής, είχε ως γνώμονα την προστασία των προσωπικών δεδομένων των πελατών της. Με προσεκτική επεξεργασία πρόσφερε τις απαραίτητες πληροφορίες, χωρίς να προκύπτει από πουθενά η δυνατότητα ταυτοποίησης των πελατών. Για τους παραπάνω λόγους θα διατηρηθεί η ανωνυμία της.

Η επεξεργασία των δεδομένων που θα χρησιμοποιηθούν και η εφαρμογή τεχνικών Data Mining, θα συνδράμουν στην κατανόηση των συμπεριφορών των πελατών και στη συνέχεια στην εφαρμογή των κατάλληλων στρατηγικών Marketing. Η επίτευξη του σκοπού θα επιτευχθεί με την υλοποίηση δύο τεχνικών Data Mining: του RFM και του Market Basket Analysis, οι οποίες θα πραγματοποιηθούν μέσω των εφαρμογών SAS Enterprise Guide και SAS Enterprise Miner. Με τη χρήση του RFM μοντέλου το λογισμικό θα τμηματοποιήσει τους πελάτες ανάλογα με τα κοινά χαρακτηριστικά που τους διακρίνουν. Οι πληροφορίες από κάθε πελάτη που θα χρησιμοποιηθούν είναι η ημερομηνία τελευταίας αγοράς (Recency), ο αριθμός των παραγγελιών που έχει πραγματοποιήσει στο κατάστημα (Frequency) και το σύνολο της αξίας των αγορών (Monetary). Αυτή η μέθοδος θα απαντήσει σε ερωτήματα σχετικά με τις ομάδες πελατών που είτε έχουν αξία για την επιχείρηση, είτε μπορεί να φέρουν αξία στο μέλλον. Στην συνέχεια, το dataset των συναλλαγών θα αναλυθεί με την τεχνική Market Basket Analysis για να παραχθούν συσχετίσεις στα προϊόντα

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

που αγοράζουν οι πελάτες. Θα δοθούν απαντήσεις σε ερωτήματα όπως ποια προϊόντα συνδυάζονται στις αγορές και ποιες προωθητικές ενέργειες θα μπορούσαν να χρησιμοποιηθούν την πώληση προϊόντων που δε διαφημίζονται σωστά.

1.3.1 Συνεισφορά

Η συνεισφορά της διπλωματικής συνοψίζεται ως εξής:

1. Μελετήθηκε η λειτουργία ενός ηλεκτρονικού καταστήματος στην Ελλάδα
2. Αναλύθηκαν τα στοιχεία πωλήσεων
3. Υλοποιήθηκαν δύο τεχνικές ανάλυσης συμπεριφορών των πελάτων
4. Μέσω του RFM ολοκληρώθηκε το Customer Segmentation
5. Μέσω του Market Basket Analysis επιτεύχθηκε η παραγωγή κανόνων συσχέτισης σε συναλλαγές πελατών
6. Από τα αποτελέσματα των αλγορίθμων εξήχθησαν σημαντικά συμπεράσματα, τα οποία μετατράπηκαν σε επιχειρηματικές προτάσεις στον τομέα του Marketing προς την επιχείρηση

1.4 Οργάνωση κειμένου

Τα κεφάλαια του εγγράφου περιγράφονται ως εξής:

1. Εισαγωγή – Αναφορά στα Big Data και τη λειτουργία των E-shops
2. Σχετικές εργασίες – Αναφορά σε θεματικές εργασίες που ενέπνευσαν το θέμα της διπλωματικής
3. Θεωρητικό Υπόβαθρο – Καταγράφεται η θεωρία, η οποία θα χρησιμοποιηθεί για την υλοποίηση των τεχνικών data mining
4. Πως οι τεχνικές data analytics συμβάλουν στην αύξηση των εσόδων των ηλεκτρονικών καταστημάτων - Ορισμός του προβλήματος που θα απαντηθεί
5. Η Επιχείρηση – Περιγραφή και ανάλυση των στοιχείων της υπό εξέτασης επιχείρησης
6. Εφαρμογή RFM Analysis και Market Basket Analysis και Αξιολόγηση αποτελεσμάτων – Εφαρμογή των τεχνικών και αξιολόγηση των αποτελεσμάτων
7. Τεχνικές Λεπτομέρειες – Επεξηγούνται τεχνικές ενέργειες που πραγματοποιήθηκαν για την υλοποίηση των αναλύσεων
8. Επίλογος – Σύνοψη των αποτελεσμάτων και καταγραφή των συμπερασμάτων

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

2

Σχετικές εργασίες

Οι θεματικές περιοχές αφορούν τις εφαρμογές Data mining technique Association Rules και Customer Segmentation σε επιχειρήσεις προκειμένου να παραχθούν αποτελέσματα και να προκύψει όφελος από την χρήση αυτών.

2.1 Customer Segmentation - Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining

Μέσω του άρθου οι (Chen, Sain, & Guo, 2012) αποδεικνύουν πως μπορούν οι μικρές επιχειρήσεις να γνωρίσουν τους υφιστάμενους πελάτες τους και να εφαρμόσουν πελατοκεντρικές στρατηγικές Marketing πιο αποτελεσματικά με την εφαρμογή της τεχνικής data mining RFM Analysis

2.2 Association Rules - Finding Ideal Menu Items Assortments: An Empirical Application of Market Basket Analysis

Οι (Ting, Pan, & Chou, 2010) μέσω του άρθρου τους εφαρμόζουν την τεχνική Market Basket Analysis σε παραγγελίες γευμάτων σε εστιατόρια προκειμένου να προκύψουν προτάσεις γευμάτων και συνδυασμοί αυτών σε πελάτες που δεν έχουν αποφασίσει τι επιθυμούν.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

3

Θεωρητικό υπόβαθρο

Το Marketing, όπως το γνωρίζουμε σήμερα, έχει παρέλθει από πάρα πολλές μεταβολές στο χρόνο ύπαρξής του. Στην πραγματικότητα, η ιστορία του ξεκινά στο μακρινό 1500 π.Χ., όταν οι λαοί στην περιοχή της Μεσοποταμίας ξεκινήσαν το εμπόριο. Ο χρονικός προσδιορισμός υποστηρίζεται από αρκετούς ιστορικούς, καθώς έχουν καταλήξει στο ότι οι έμποροι αποτύπωναν στα προϊόντα τους ένα είδος λογότυπου, που σηματοδοτούσαν στους αγοραστές την προέλευση των προϊόντων. Η αλματώδης ανάπτυξη του Marketing συμβαδίζει με την έκρηξη νέων τεχνολογιών, οι οποίες έδωσαν τη δυνατότητα να παραχθούν ακόμα περισσότερα δεδομένα προς επεξεργασία και ανάλυση. Οι παραδοσιακές καμπάνιες Marketing έχουν ως στόχο να χρησιμοποιούν πληροφορίες και να αναγνωρίζουν πελάτες ως υποψήφιους για να χτίσουν μια συνεχή σχέση μαζί τους για το μεγαλύτερο όφελος των πελατών και, φυσικά, για το μεγαλύτερο κέρδος της εταιρείας. Σε αυτές τις καμπάνιες προωθείται το ίδιο μήνυμα, το ίδιο offer χρησιμοποιώντας το ίδιο μέσο απευθυνόμενοι σε μεγάλο πληθυσμό (Kahan, 1998). Οι παλαιότερες επιχειρήσεις θεωρητικά είναι οι πρώτες που χρησιμοποίησαν στοχευμένες στρατηγικές Marketing. Ο κάθε έμπορος γνώριζε τους πελάτες του, συνήθως εκπαιδευόταν στις συνήθειές τους και, κατά συνέπεια, γνώριζε τις προτιμήσεις τους. Όπως τονίζει ο (Kahan, 1998) στο άρθρο του, η ανθρωπότητα, παρά την ελευθερία σκέψης, συνηθίζει να επαναλαμβάνει συμπεριφορές και συνήθειες, με άμεση συνέπεια η συμπεριφορά των ανθρώπων να είναι προβλέψιμη. Έτσι, το μειονέκτημα των πελατών γίνεται πλεονέκτημα των επιχειρήσεων, όπου με την εκμετάλλευσή τους επιτυγχάνονται καλύτερα επιχειρησιακά αποτελέσματα. Στη σύγχρονη εποχή, η επεξεργασία δεδομένων είναι απαραίτητη προκείμενου να είναι εφικτή η εκμετάλλευσή τους για τη λήψη αποφάσεων εντός στρατηγικών μεθόδων ενός επιχειρησιακού οργανισμού.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

3.1 Digital Marketing Strategies

Η έννοια του Digital Marketing δημιουργήθηκε από το Internet και τις μηχανές αναζήτησης. Πιο συγκεκριμένα, είναι κάτι περισσότερο από ένα απλό Marketing διαδικτύου ή ηλεκτρονικό Marketing. Πλέον αποτελεί μια σύγχρονη δραστηριότητα Marketing, που περιλαμβάνει τη χρήση διαφορετικών μέσων που βασίζονται στο Internet, όπως email, ιστοσελίδες, ιστολόγια ή social media (Forghani, Sheikh, Hosseini, & Sana, 2021).

Μέσω του Digital Marketing είναι εφικτή η άμεση αλληλεπίδραση των εταιριών με τους πελάτες. Εντοπίζονται πολλές ευκαιρίες για προωθητικές ενέργειες και η ανατροφοδότηση είναι άμεση σχετικά με τις αντιδράσεις των πελατών. Οι πληροφορίες αυτές επιτρέπουν στις επιχειρήσεις να βελτιώσουν τις ενέργειές τους και με αυτό τον τρόπο να πετύχουν καλύτερα αποτελέσματα. Η εφαρμογή τους μέσω ψηφιακής πλατφόρμας επιτρέπει την επικοινωνία με μεγαλύτερο κοινό.

Πλεονεκτήματα	**Μειονεκτήματα**
Εύκολη επικοινωνία	Απαιτεί εξειδικευμένο προσωπικό
Χαμηλό κόστος	Απαιτεί υψηλή τεχνολογία
Εξατομίκευση	Online μπορεί ένας χρήστης να σχολιάσει αρνητικά τις προωθητικές ενέργειες. Αυτό ελλοχεύει τον κίνδυνο να δημιουργήσει αρνητική φήμη για την εταιρία
Δεν υπάρχει χρονικός περιορισμός διεξαγωγής	΄Ελλειψη Εμπιστοσύνης πελάτη, Δυσπιστία λόγω συχνών φαινομένων ψηφιακής απάτης
Εύκολη Αξιολόγηση αποτελεσμάτων	

Πίνακας 1 - Πλεονεκτήματα και Μειονεκτήματα των Digital Στρατηγικών

3.1.1 Στρατηγικές Digital Marketing

Search Engine Optimization (SEO): Κατά την αναζήτηση μιας πληροφορίας σε μηχανή αναζήτησης, οι επιχειρήσεις που έχουν υψηλή βαθμολογία εμφανίζονται στα πρώτα αποτελέσματα αυτής. Με αυτόν τον τρόπο είναι πιο πιθανό οι ενδιαφερόμενοι να αγοράσουν από μία επιχείρηση με υψηλή βαθμολογία.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Search Engine Marketing (SEM): Αφορά διαφημίσεις σε ιστοσελίδες με μεγάλη επισκεψιμότητα, όπου οι επιχειρήσεις μπορούν να χρησιμοποιήσουν μοντέλα όπως πληρωμή ανά click, αγορά clicks (δηλαδή η διαφήμιση εμφανίζεται για συγκεκριμένα clicks πελατών).

Social Media Marketing (SMM): Η χρήση των μέσων κοινωνικής δικτύωσης είναι τεράστια τα τελευταία χρόνια. Είναι τα μέσα με τα οποία οι άνθρωποι επικοινωνούν μεταξύ τους και μεταφέρουν πληροφορίες και εικόνες από τη ζωή τους. Οι επιχειρήσεις μπορούν να διαφημίζουν τα προϊόντα τους σε αυτά και με αυτόν τον τρόπο έχουν πρόσβαση σε κοινότητες που δεν θα είχαν με το παραδοσιακό Marketing.

Content Marketing: Πρόκειται για πρωτοποριακή στρατηγική, όπου δημοσιεύονται κείμενα και αξίες της επιχείρησης προσπαθώντας να δημιουργηθεί «καλή» εικόνα προς τους πελάτες. Χαρακτηριστικό παράδειγμα η καμπάνια "Our Food, Your Questions" της McDonald's στον Καναδά τον Οκτώβριο του 2014. Μέσα από την επιτυχημένη καμπάνια προσπάθησαν να αλλάξουν κακές φήμες σχετικά με την συσκευασία των προϊόντων, την προετοιμασία των γευμάτων, την ποιότητα του φαγητού κ.α., οι οποίες είχαν δημιουργηθεί από αρνητικά σχόλια στο διαδίκτυο (Jarboe, 2015).

Recommender Engine (RE): Μέσω αυτή της στρατηγικής προτείνονται στον καταναλωτή πιθανά προϊόντα ή υπηρεσίες για αγορά. Οι πληροφορίες αντλούνται από διάφορες πηγές που επισκέφθηκε ο χρήστης στο διαδίκτυο για αγορές προϊόντων, αναζήτηση βιβλίων/ταινιών κτλ.

Email Marketing: Οι διαφημιστικές ενέργειες που προωθούνται μέσω ηλεκτρονικής αλληλογραφίας, ονομάζονται Email Marketing. Πρόκειται για μια οικονομική στρατηγική, καθώς δεν απαιτούνται προϊόντα υψηλής τεχνολογίας και ο χρόνος προετοιμασίας είναι μικρός.

3.2 Data exploration

Για την παραγωγή χρήσιμων και ασφαλών αποτελεσμάτων μέσω της εκμετάλλευσης των πληροφοριών που μας περιτριγυρίζουν, απαιτείται μεγάλος όγκος δεδομένων. Όσο μεγαλύτερο όγκο δεδομένων διαθέτουμε, τόσο το καλύτερο (Analytics in a Big Data World. The Essential Guide to Data Science and its Applications by Bart Baesens).

Δε ζούμε όμως σε έναν ιδανικό κόσμο και το ίδιο ισχύει και για τα δεδομένα. Η ποιότητα τους αποτελεί έναν από τους σημαντικότερους παράγοντες στα Data Analytics. Αυτά μπορεί να προέρχονται από διασκορπισμένες πληροφορίες στον παγκόσμιο ιστό, σε βάσεις δεδομένων και σε

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

άλλες πηγές. Η προσεκτική επεξεργασία τους συντελεί στην παραγωγή καλύτερων αποτελεσμάτων και ασφαλέστερων συμπερασμάτων.

Οι συναλλαγές αποτελούν την πιο βασική και πρακτική πηγή δεδομένων, διότι εντοπίζονται με δομημένο τρόπο και με πολλές λεπτομέρειες στα στοιχεία συναλλαγής, όπως το κόστος, ο τρόπος πληρωμής, κτλ. Παράλληλα, διατηρούνται πληροφορίες, όπως το χρονικό διάστημα με αποτέλεσμα να είναι εφικτή η ιστορική αναζήτηση. Οι συγκεκριμένες πληροφορίες, καθώς υφίστανται σε μεγάλους πίνακες δεδομένων, αποθηκεύονται σε βάσεις δεδομένων που ονομάζονται Online transaction processing (OLTP).

Μερικά παραδείγματα OLTP:

- ATMs
- Online banking
- Online airline ticket booking
- Sending a text message
- Order entry
- Add a book to shopping card

Η κύρια διαφορά των OLTP με τις Online Analytical Processing (OLAP) βάσεις δεδομένων είναι ότι στη δεύτερη κατηγορία χρησιμοποιούνται εργαλεία για την ανάλυση των δεδομένων, ενώ στις OLTP εφαρμόζονται εργαλεία με κύρια δραστηριότητα την υποστήριξη των συναλλαγών.

Έντονο ενδιαφέρον ανάλυσης παρουσιάζουν επίσης τα αδόμητα δεδομένα που εντοπίζονται σε κείμενα όπως emails, Internet sites, text messages, κτλ. Όμως, καθώς η επεξεργασία τους απαιτεί επιπλέον εργαλεία και μεγαλύτερη εξειδίκευση στην χρήση τους, το κόστος είναι ακόμα μεγαλύτερο. Τέλος, μια επιπλέον πηγή δεδομένων είναι αυτά που προέρχονται από ειδικούς εμπειρογνώμονες, οι οποίοι κατέχουν σημαντικές εξειδικευμένες γνώσεις σε έναν συγκεκριμένο τομέα (οικονομικός/χρηματοοικονομικός τομέας, τραπεζικά δάνεια, κτλ). Τα δεδομένα αυτά χαρακτηρίζονται, συνεπώς, πιο ποιοτικά. Σε μια έρευνα ανάλυσης, η τεχνογνωσία είναι πολύ σημαντική, καθώς προσθέτει επιχειρηματική εμπειρία και καθοδηγεί τη μοντελοποίηση προς την σωστή κατεύθυνση και η ερμηνεία των αποτελεσμάτων θα πραγματοποιηθεί με τη σωστή προοπτική. Παραδείγματα τέτοιων εταιριών είναι οι Dun & Bradstreet και Thomson Reuters, οι οποίες συλλέγουν δεδομένα από έναν συγκεκριμένο χώρο και στη συνέχεια κατασκευάζουν μοντέλα, το οποία είτε πωλούνται σε ενδιαφερόμενους πελάτες, είτε παρέχουν υπηρεσίες με την παραγωγή τους. Η πιστωτική βαθμολογία FICO στις ΗΠΑ είναι η πιο διαδεδομένη.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

3.2.1 Δειγματοληψία

Όπως αναφέραμε παραπάνω τα δεδομένα βρίσκονται παντού γύρω μας. Μπορούμε να επεξεργαστούμε στοιχεία από έναν ολόκληρο πληθυσμό ή να εργαστούμε πάνω σε ένα δείγμα του. Το πιο σύνηθες είναι να έχουμε διαθέσιμα δεδομένα από ένα δείγμα πληθυσμού. Ένα πελατολόγιο μιας εταιρίας είναι ολόκληρος ο πληθυσμός της, ενώ ταυτόχρονα μπορεί να θεωρηθεί υποσύνολο, άρα και δείγμα, ενός συνόλου καταναλωτών αγοραστικής δύναμης ενός συγκεκριμένου κλάδου.

Η δειγματοληψία προκύπτει από διάφορους τύπους. Η διάκρισή τους είναι η παρακάτω:

- **Δειγματοληψία πιθανοτήτων:**

Σε αυτή τη μεθοδολογία κάθε πιθανότητα, η οποία είναι διάφορη του μηδενός, μπορεί να επιλεγεί στο δείγμα και ο προσδιορισμός της είναι ακριβής.

- **Δειγματοληψία μη πιθανοτήτων:**

Πραγματοποιείται στις περιπτώσεις που δεν είναι εφικτή η δειγματοληψία με πιθανότητα ή όταν είναι αναγκαίο να πραγματοποιηθεί γρήγορα η εφαρμογή της έρευνας όπως μια πιλοτική (Baker, 1988)

Στις περισσότερες εφαρμογές επιλέγεται η δειγματοληψία πιθανοτήτων με τις παρακάτω διακρίσεις:

- **Απλή τυχαία δειγματοληψία:** Κάθε στοιχείο του πληθυσμού έχει την ίδια πιθανότητα να επιλεγεί. Αντιπροσωπευτικό παράδειγμα είναι η κλήρωση τυχερών αριθμών. Η χρήση της ενδέχεται ωστόσο να αφήσει περιοχές του πληθυσμού χωρίς κάλυψη.
- **Συστηματική δειγματοληψία:** Εάν το δείγμα είναι καταγεγραμμένο σε λίστα, μπορούμε να εφαρμόσουμε συστηματική δειγματοληψία. Με την ορθή εφαρμογή της επιτυγχάνεται η επιλογή αντιπροσωπευτικού δείγματος.
- **Στρωματοποιημένη δειγματοληψία:** Ο πληθυσμός χωρίζεται σε στρώματα και στη συνέχεια με απλή τυχαία δειγματοληψία επιλέγονται επιμέρους τμήματα.
- **Δειγματοληψία σε συστάδες:** Ο πληθυσμός χωρίζεται σε ομάδες και στη συνέχεια επιλέγεται δείγμα από τις ομάδες. Χρησιμοποιείται κυρίως σε έρευνες γεωγραφικών περιοχών.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

3.2.2 Κατηγορίες δεδομένων

Προκειμένου να μπορέσουμε να αναλύσουμε τα διαθέσιμα δεδομένα, θα πρέπει να είμαστε σε θέση να γνωρίζουμε τους διαφορετικούς τύπους που τα διακρίνουν. Ανάλογα την μορφή που τηρούν, μπορούμε να τα επεξεργαστούμε, να τα μεταβάλουμε, να τα αντιστοιχίσουμε και να τα ταξινομήσουμε. Συνεπώς, προκύπτει η παρακάτω διάκριση (Sharma, 2020):

- **Nominal**: Πρόκειται για κατηγορία δεδομένων μη αριθμητικών, άρα και χωρίς διάταξη. Παράδειγμα είναι το χρώμα αντικειμένων. Δηλαδή δεν μπορεί το κίτρινο να συγκριθεί αριθμητικά με το κόκκινο.
- **Ordinal**: Οι συγκεκριμένες τιμές έχουν μια φυσική σειρά, αλλά και κατηγορία διάταξης, δηλαδή τοποθετούνται σε κάποιου είδους σειρά ανάλογα με τη θέση τους σε κλίμακα. Αυτή είναι και η κύρια διαφορά με τις Nominal, αλλά παρόλα αυτά δεν μπορούν να πραγματοποιηθούν αριθμητικές πράξεις μεταξύ τους. Παράδειγμα είναι η βαθμολογία γραπτών εξετάσεων με τιμές Α, Β, Γ και Δ.
- **Discrete**: Περιλαμβάνονται μόνο ακέραιοι αριθμοί ως τιμές και δεν υποδιαιρούνται σε περισσότερα μέρη. Παράδειγμα είναι ο αριθμός θεατών σε ένα θέατρο.
- **Continuous**: Τα συνεχή δεδομένα ορίζονται ως μεταβλητές οι οποίες μπορούν να λάβουν οποιαδήποτε τιμή από ένα εύρος τιμών. Τέτοια παραδείγματα είναι η θερμοκρασία ή ο χρόνος απόκρισης μιας ομάδας.

3.2.3 Data Cleansing

Η διαδικασία Data Cleansing είναι απαραίτητη για τη δημιουργία ποιοτικών δεδομένων προς ανάλυση και αυτό προκύπτει όσο μεγαλύτερο είναι το μέγεθος του dataset. Οι πιο συχνές περιπτώσεις είναι εκείνες, όπου λείπουν στοιχεία (Missing Values). Αυτό μπορεί να συμβεί για διάφορους λόγους, όπως να μην έχουν καταχωρηθεί σκόπιμα (παράδειγμα ο χρήστης δεν επιθυμεί να δηλώσει την ημερομηνία γέννησης) ή να έχουν συμπληρωθεί με τυπογραφικά λάθη ή να έχουν χαθεί κατά την διαδικασία μιας συγχώνευσης πινάκων ή να έχουν καταχωρηθεί διπλότυπες παρατηρήσεις. Οι πιο συχνές μέθοδοι Data Cleansing είναι οι παρακάτω (Little & Rubin, 2020):

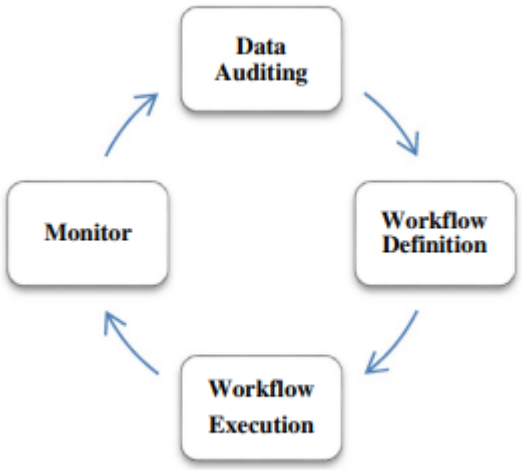
- **Διαγραφή:** Η πιο απλή διαδικασία όπου διαγράφονται παρατηρήσεις όταν απουσιάζουν τα στοιχεία. Μπορούν να διαγραφούν βάσει λίστας ή κατά ζεύγη.



- **Αντικατάσταση:** Αντί να διαγραφούν οι τιμές, αντικαθίστανται με τιμές μέσων άνευ όρων ή υπό όρους ή εναλλακτικά αντικαθίστανται με τιμές από παρόμοιες περιπτώσεις όπου οι παρατηρήσεις είναι πλήρεις.

Υπάρχουν περιπτώσεις όπου η διατήρηση των Missing Values έχει νόημα, διότι ο χρήστης μπορεί εσκεμμένα να διατήρησε κενή την πληροφορία. Για παράδειγμα, δεν ήθελε να αποκαλύψει το ετήσιο εισόδημά του και κατά συνέπεια άφησε κενό το εν λόγω πεδίο συμπλήρωσης.

Κατά τους Müller & Freytag μια προοπτική διαδικασίας Data Cleansing είναι μια σειρά από προκαθορισμένες ενέργειες που απαιτούνται για να καταλήξουμε σε έναν όγκο δεδομένων που θα είναι ακριβής και όσο το δυνατό πιο αντιπροσωπευτικό το δείγμα του πληθυσμού. Περιλαμβάνει τα παρακάτω τέσσερα κύρια βήματα: Data Auditing, Workflow Specification, Workflow Execution και Post-Processing & Controlling.



Εικόνα 2 - Data Cleansing Flow (Müller & Freytag, 2005)

3.3 Data Mining

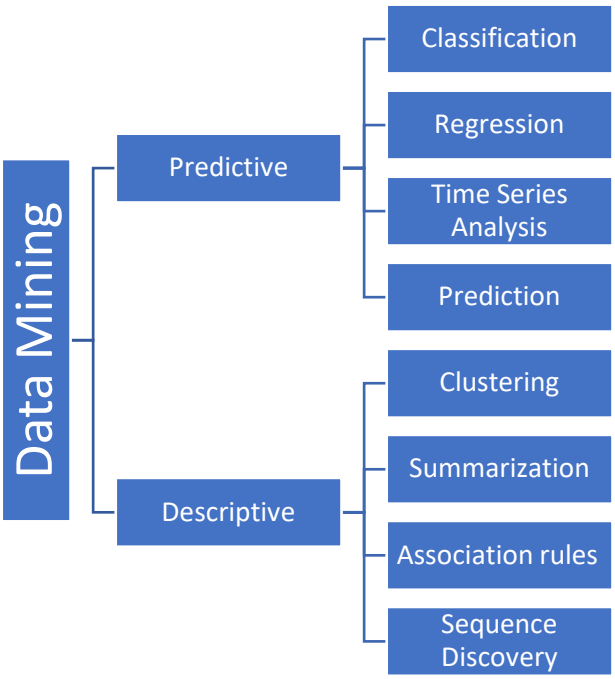
Ο ορισμός του Data Mining αφορά σε όλες εκείνες τις εργασίες που πραγματοποιούνται προκειμένου να εξαχθούν αποτελέσματα από μεγάλο όγκο δεδομένων. Αυτές οι εργασίες περιλαμβάνουν πολλούς διαφορετικούς αλγορίθμους, ώστε να ολοκληρώνουν διάφορα Tasks σκοπεύοντας να προσαρμόσουν ένα μοντέλο στα δεδομένα. Η δομή τους αποτελείται από τρία μέρη,

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

το model, το preference και το search. Το μοντέλο που δημιουργείται είναι είτε Predictive, είτε Descriptive. Ένα Predictive μοντέλο μπορεί να χρησιμοποιεί ιστορικά στοιχεία για να προβλέψει. Αντιθέτως, το Descriptive μοντέλο προσδιορίζει μοτίβα και σχέσεις στα δεδομένα και χρησιμεύει να στο να διερευνηθούν στοιχεία στα υφιστάμενα δεδομένα, ενώ από την άλλη πλευρά το Predictive προβλέπει νέες ιδιότητες. (Dunham, 2002).

Παρακάτω παρατίθενται συνοπτικά οι Data Mining λειτουργικότητες.

1. **Classification:** Συγκεντρώνει τα δεδομένα σε ομάδες ή κλάσεις.
2. **Regression:** Εντάσσει τα δεδομένα σε μια μεταβλητή πρόβλεψης πραγματικής αξίας.
3. **Time Series Analysis**: Αξιολογεί τα στοιχεία σε περίοδο χρόνου.
4. **Prediction:** προβλέπει το αποτέλεσμα ταξινομώντας τα διαθέσιμα στοιχεία.
5. **Clustering:** Παρόμοια λειτουργικότητα με το Classification, αλλά εδώ οι ομάδες δεν είναι προκαθορισμένες.
6. **Summarization:** Ομαδοποιεί τα δεδομένα σε υποσύνολα με απλά χαρακτηριστικά.
7. **Association rules:** Αξιολογεί και δημιουργεί συσχετίσεις μεταξύ των δεδομένων.
8. **Sequence Discovery:** Προσδιορίζει διαδοχικά μοτίβα σε δεδομένα.



Εικόνα 3 - Data Mining Models and Tasks Dunham, 2002)

Στο πλαίσιο της εκπόνηση της μεταπτυχιακής διατριβής θα αναλυθούν παρακάτω οι λειτουργικότητες Clustering και Association rules.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

3.4 RFM Model (Recency, Frequency, and Monetary)

Η πιο συχνά χρησιμοποιούμενη κατηγορία τμηματοποίησης είναι η συμπεριφορική (behavioural segmentation) και η τεχνική RFM segmentaion (παράγεται από τα ακρωνύμια Recency, Frequency και Monetary value), η οποία σχετίζεται με συμπεριφορικά δεδομένα. Το 1995 οι Bult και Wansbeek δημοσιεύσαν ένα άρθρο για το Direct Marketing, όπου περιγράφουν το RFM ως τεχνική επιλογής πελατών για προωθητικές ενέργειες. Το επονομαζόμενο μοντέλο RFM αποτελεί μια απλή μέθοδο που χωρίζει κάθε μεταβλητή RFM σε κατηγορίες (υποκειμενικά επιλεγμένες από τον ερευνητή) και εκχωρεί πιθανότητες σε κάθε κατηγορία κάθε χαρακτηριστικού σύμφωνα με τη συμπεριφορά της διαφορετικής απόκρισης (Bult & Wansbeek, 1995). Το Recency υπολογίζει το διάστημα από την τελευταία φορά που ο πελάτης ανταποκρίθηκε σε μια προωθητική ενέργεια και αγόρασε αγαθά ή απλά επισκέφθηκε ένα κατάστημα και αγόρασε προϊόντα ή υπηρεσίες. Το διάστημα αυτό μπορεί να αποτυπωθεί σε μήνες, εβδομάδες ή μέρες. Το Frequency υπολογίζει το πλήθος των αγορών που έχει πραγματοποιήσει ο ίδιος πελάτης στο ίδιο κατάστημα σε ένα συγκεκριμένο χρονικό διάστημα. Το Monetary αθροίζει το κόστος των αγορών που πραγματοποίησε ο πελάτης σε ένα συγκεκριμένο χρονικό διάστημα.

Το μοντέλο RFM συνδράμει στο να γνωρίζουν καλύτερα οι επιχειρήσεις τους πελάτες τους και να σχεδιάσουν ακόμα πιο στοχευμένα τις στρατηγικές Marketing. Η προσέλκυση των πελατών αποτελεί μια απαραίτητη ενέργεια για κάθε εμπορικό οργανισμό, καθώς το προς πώληση προϊόν/υπηρεσία δεν είναι πάντα ελκυστικό από μόνο του. Το σύγχρονο Customer Relationship Management απαιτεί να διατηρεί μια επιχείρηση το πελατολόγιο της γνωρίζοντας την αξία των πελατών της (Wei, Lin, & Wu, 2010). Για την ολοκλήρωσή του χρησιμοποιείται ο αλγόριθμος k-means με αποτέλεσμα οι πελάτες να τμηματοποιούνται σε ομάδες. Έτσι παρέχεται η δυνατότητα στις επιχειρήσεις με ηλεκτρονικό κατάστημα να αντιμετωπίζουν κάθε πελάτη ως άτομο με εξατομικευμένη κατανόηση και να βασίζονται στην επιχειρηματική ευφυΐα με επίκεντρο τον πελάτη (Chen, Sain, & Guo, 2012). Για πολλές εταιρίες η τμηματοποίηση πελατών αποτελεί πλέον μια συνήθης πρακτική και αναγκαία διαδικασία για την υιοθέτηση πελατοκεντρικών ενεργειών. Η χρήση παρόμοιων εργαλείων δίνει το ανταγωνιστικό πλεονέκτημα στην αγορά έναντι των άλλων επιχειρήσεων, διότι μπορεί να αυξήσει το πελατολόγιο της, να αποδώσει μεγαλύτερο κέρδος και να μειώσει τα κόστη των παραγγελιών. Η ανάλυση του μοντέλου εξετάζει την ιστορικότητα συναλλαγών των πελατών, τη συχνότητα παραγγελιών, το χρονικό διάστημα τους και τα έσοδα που έχουν αποφέρει στην επιχείρηση, ώστε να χαρτογραφήσει τους πελάτες που της δίνουν αξία.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Καθώς αποτελεί ένα από τα πιο χρησιμοποιούμενα μοντέλα, το RFM, όπως περιγράψαμε, παράγει ένα dataset με τρεις μετρήσεις. Από τις τρεις εν λόγω διαστάσεις, το Recency διαπιστώνεται να είναι το πιο σημαντικό, αλλά πάντα επηρεάζονται και τα τρία από τη φύση των προϊόντων/υπηρεσιών της επιχείρησης. Σύμφωνα με έρευνες που έχουν πραγματοποιηθεί έχει αποδειχθεί ότι, όσο πιο πρόσφατα έχει ολοκληρώσει ένας πελάτης μια συναλλαγή με την εταιρία, τόσο πιο πιθανό είναι να συνεχίσει να κάνει συναλλαγές και στο μέλλον. Άρα λοιπόν ένας «καλός» πελάτης είναι αυτός που έχει μικρό Recency.

Το Frequency αφορά στο πόσες φορές έχει αλληλοεπιδράσει ο πελάτης με τον οργανισμό για μια συγκεκριμένη χρονική περίοδο ανάλυσης των δεδομένων της εταιρίας,. Για παράδειγμα, πόσες φορές έχει ψωνίσει ένας πελάτης από ένα ηλεκτρονικό κατάστημα. Όσο πιο συχνά αγοράζει κάποιος πελάτης από την εταιρία, τόσο πιο πιθανό είναι να συνεχίσει να αγοράζει και στο μέλλον. Συνακόλουθα, «καλός» πελάτης χαρακτηρίζεται ένας πελάτης με μεγάλο Frequency.

Τέλος, το Monetary είναι το συνολικό ποσό των χρημάτων που έχει ξοδέψει ο πελάτης στην εταιρία ή ο μέσος όρος ανά συναλλαγή για μια συγκεκριμένη χρονική περίοδο. Αυτό χρησιμοποιείται υπό την υπόθεση, η οποία βέβαια έχει αποδειχθεί, ότι όσο πιο πολλά χρήματα έχει ξοδέψει ένας πελάτης στο παρελθόν, τόσο πιο πιθανό είναι να ξοδέψει και στο μέλλον. Υψηλό monetary σημαίνει πελάτης που ξοδεύει χρήματα στο κατάστημα, άρα είναι «καλός» πελάτης.

Συνοψίζοντας, ένας «καλός» πελάτης θα πρέπει να έχει χαμηλό Recency, υψηλό Frequency και υψηλό Monetary, ενώ αντίθετα «κακός» πελάτης χαρακτηρίζεται αυτός με υψηλό Recency, χαμηλό Frequency και χαμηλό Monetary. Ενδιάμεσα βέβαια υπάρχουν και άλλες κατηγορίες πελατών. Η κατηγοριοποίηση που προκύπτει δημιουργεί groups με διαφορετικές ιδιότητες και ανάλογα ονοματίζονται. Για παράδειγμα τα Segments μπορούν να ονομαστούν Best, Valuable, Churners, First Time Great Potential, First Time Uncertain και Worst.

Το μεγαλύτερο πλεονέκτημα του RFM μοντέλου είναι η λογική σύνδεση των αποτελεσμάτων, εφόσον οι μεταβλητές που έχουμε χρησιμοποιήσει είναι αντικειμενικές και παρατηρήσιμες. Παραδοσιακά η υιοθέτηση του πραγματοποιείται με την ταξινόμηση των δεδομένων βάσει των μεταβλητών RFM και στη συνέχεια διαιρούνται σε Clusters. Η ταξινόμηση πραγματοποιείται πρώτα με βάση το Recency και συγκεκριμένα με τους πελάτες που έχουν πραγματοποιήσει πιο πρόσφατα συναλλαγές, στη συνέχεια ακολουθεί η ταξινόμηση με το Frequency και τέλος με το Monetary με φθίνουσα σειρά.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

3.4.1 Πλεονεκτήματα και μειονεκτήματα χρήσης RFM

Η τμηματοποίηση RFM, όπως όλα τα μοντέλα στο Marketing, παρουσιάζει πλεονεκτήματα και μειονεκτήματα. Για να γίνει πιο κατανοητό παρακάτω καταγράφονται τα κυριότερα πλεονεκτήματα και μειονεκτήματα:

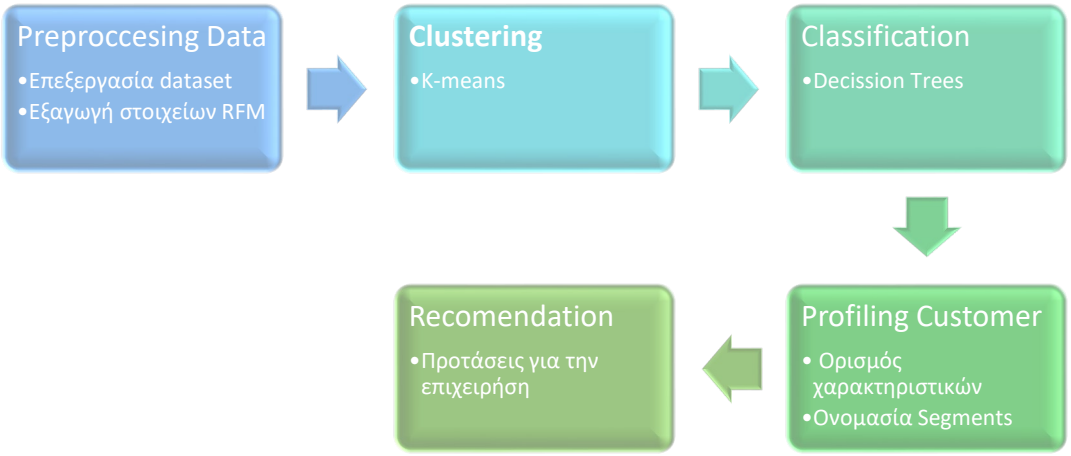
Πλεονεκτήματα	**Μειονεκτήματα**
Το κόστος απόκτησης των δεδομένων είναι πολύ χαμηλό, συνεπώς cost effective για τους ερευνητές.	Το μοντέλο είναι πολύ απλό, καθώς χρησιμοποιεί μόνο τρεις μεταβλητές, γεγονός το οποίο χαρακτηρίζεται κάποιες φορές ως μειονέκτημα.
Η ποσοτικοποίηση της συμπεριφοράς των πελατών με τις 3 μεταβλητές είναι εξαιρετικά εύκολη.	Σχεδόν όλα τα μοντέλα στο Marketing υποθέτουν στασιμότητα της συμπεριφοράς των πελατών. Έτσι και το RFM υποθέτει ότι μια συμπεριφορά των πελατών στο παρελθόν, θα συνεχίσει να είναι η ίδια και στο μέλλον
Το πιο σημαντικό είναι ότι είναι πολύ εύκολα ερμηνεύσιμο και εξηγήσιμο και επομένως, μπορεί να «πουληθεί» η ανάλυση αυτή σε ανθρώπους, οι οποίοι δεν έχουν εξειδικευμένες γνώσεις στατιστικής και ποσοτικών μεθόδων, π.χ. στη διοίκηση ή στους επιχειρησιακούς χρήστες.	- Υπάρχει το Παραδοσιακό και το Μοντέρνο μοντέλο RFM segmentaion. Το παραδοσιακό εστιάζει περισσότερο στους καλύτερους πελάτες και όχι στους χειρότερους, αλλά πολλές φορές μπορούμε να χρησιμοποιήσουμε και τους χειρότερους πελάτες για να τους «τροποποιήσουμε» και να τους κάνουμε καλούς πελάτες.

Πίνακας 2 - Πλεονεκτήματα και Μειονεκτήματα RFM

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

3.4.2 RFM Μεθοδολογία

Για την εφαρμογή της έρευνας απαιτείται η ολοκλήρωση αρκετών σταδίων, όπως η λήψη δεδομένων, επεξεργασία αυτών, ομαδοποίηση, ταξινόμηση και τέλος το customer profiling.



Εικόνα 4 - RFM Flow

Preparing and Pre-processing Data

Σε αυτό το στάδιο συγκεντρώνονται όλες οι απαραίτητες πληροφορίες που θα στοιχειοθετήσουν την RFM ανάλυση. Για την επιλογή των ορθών δεδομένων απαιτούνται πληροφορίες όπως το κόστος, ο χρόνος και η συχνότητα. Στη συνέχεια είναι εφικτή η παραγωγή του dataset με τα γνωστά metrics Recency, Frequency και Monetary.

Clustering with k-Means Method

Αφού το RFM dataset είναι έτοιμο προς ανάλυση, χρησιμοποιείται ο k-Means αλγόριθμος, ο οποίος είναι μια απλή, επαναληπτική διαδικασία. Αποτελεί τον απλούστερο αλγόριθμο ομαδοποίησης και χρησιμοποιεί ως είσοδο έναν προκαθορισμένο αριθμό συστάδων που είναι το k. Το means σημαίνει ότι είναι ένας μέσος όρων στοιχείων. Αρχικά, προσδιορίζονται τα Clusters που είναι επιθυμητό να προκύψουν. Τα δεδομένα τοποθετούνται στο κοντινότερο cluster και για να γίνει αυτό πρέπει να μετρηθεί η απόσταση από τον πρώτο πελάτη στο κοντινότερο cluster. Αφού όλα τα δεδομένα τοποθετηθούν στο πλησιέστερο cluster, πρέπει να υπολογιστεί ξανά το νέο κέντρο cluster με βάση τα μέσα μέλη του cluster και επαναλαμβάνεται η διαδικασία έως ότου το κεντρικό σημείο κάθε cluster να μην αλλάζει πλέον και κανένα στοιχείο να μην μετακινείται από το ένα cluster στο άλλο.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Classification with Decision Tree Method

Το βήμα αυτό δεν είναι απαραίτητο να εκτελεστεί. Ωστόσο η χρήση του εφαρμόζεται με στόχο να προβλεφθεί η συμπεριφορά ενός πελάτη βάσει των χαρακτηριστικών RFM.

Profilling Customer

Το profiling Customer προκύπτει από την επιμέρους ερμηνεία και την κατανόηση των clusters που δημιουργήθηκαν.

Recommendations

Αφού το πελατολόγιο έχει χαρτογραφηθεί και τα χαρακτηριστικά των πελατών έχουν ερμηνευθεί, τότε είναι εφικτό να συνταχθούν οι απαραίτητες προτάσεις προς την επιχείρηση για την πελατοκεντρική προσέγγιση των πελατών της.

3.4.3 Segmentation

Η τεχνική της διαίρεσης των πελατών σε ομάδες βάσει κοινών χαρακτηριστικών είναι γνωστή ως τμηματοποίηση πελατών, η οποία επιτρέπει σε διάφορες εταιρείες να εμπορεύονται κάθε ομάδα αποτελεσματικά και κατάλληλα. Αναφορές στο Customer Segmentation για πρώτη φορά συναντάμε στον Smith (Smith, 1956). Η γενική ιδέα είναι να κατηγοριοποιηθούν οι πελάτες σε groups με όμοια χαρακτηριστικά. Η τμηματοποίηση πελατών είναι μια από τις πιο κοινές τεχνικές που χρησιμοποιούν οι επιχειρήσεις σε όλο τον κόσμο για να στοχεύσουν ορισμένους πελάτες για ένα προϊόν ή μια υπηρεσία (Sharanjit Kaur, 2021). Επιπρόσθετα, οι τεχνικές τμηματοποίησης δίνουν την δυνατότητα στους οργανισμούς να γνωρίζουν το πελατολόγιο τους και να μπορούν να δημιουργούν νέα προϊόντα και υπηρεσίες που τους αποφέρουν επιπλεόν κέρδος. Η αγορά και οι πελάτες παρουσιάζουν μεταβαλλόμενες συμπεριφορές και χαρακτηριστικά. Διακρίνονται και επηρεάζονται από στοιχεία όπως η ηλικία, το φύλο, οι αγοραστικές συνήθειες, οι προτιμήσεις της επωνυμίας, κ.α. Κυρίως όμως, η τμηματοποίηση βασίζεται στη συμπεριφορά, τις γεωγραφικές και δημογραφικές πληροφορίες των πελατών.

Καθώς οι ηλεκτρονικές αγορές δεν είναι πλέον καινοτομία στην εποχή μας, αλλά αποτελεί μια συναλλαγή ενταγμένη στην καθημερινότητα των πελάτων, η συνεχής τροφοδότηση τους με στοιχεία, διαφημίσεις και πωλήσεις, πολλές φορές δημιουργούν σύγχυση στο καταναλωτικό κοινό. Η εξατομίκευση βάσει προτιμήσεων των πελατών έρχεται να επιλύσει σε μεγάλο βαθμό αυτότο πρόβλημα.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Οι τύποι τμηματοποίησης μπορούν να διακριθούν όπως παρακάτω:

Behavioural Segmentation

Η πιο αποτελεσματική κατηγορία τμηματοποίησης και ταυτόχρονα η πιο συνήθης σε χρήση. Η συλλογή των δεδομένων είναι σχετικά εύκολη και πραγματοποιείται με βάση τη συμπεριφορά του πελάτη. Το Profile συμπεριφοράς μπορεί να προσδιοριστεί μέσω του προτύπου αγοράς, της ποσότητας του προϊόντος, της ποιότητας του προϊόντος, της χρήσης, της μάρκας του προϊόντος κ.λπ. (Sumit Koul, 2021) .

Psychographic Segmentation

Βασίζεται στον τρόπο ζωής των πελατών, τις πεποιθήσεις (θρησκευτικές, πολιτικές, κα.), την γνώμη τους, τα ενδιαφέροντα, το επάγγελμα, τις δραστηριότητες τους, κ.α.

Geographic Segmentation

Γεωγραφική είναι η τμηματοποίηση που βασίζεται σε περιοχές όπως χώρες, περιοχές, αστικά κέντρα ή επαρχία, παράκτιες ή ορεινές, κ.λπ.

Demographic Segmentation

Η δημογραφική τμηματοποίηση βασίζεται σε πληροφορίες όπως οι κοινωνικο-δημογραφικές πληροφορίες όπως ηλικία, φύλο, εκπαίδευση, επάγγελμα, οικογενειακή κατάσταση, εισόδημα κ.λπ.

Value-based Segmentation

Αποτελεί μια μέθοδο τμηματοποίησης των πελάτων βάσει της αξίας τους. Σε γενικές γραμμές προσδιορίζονται οι πελάτες που είναι διατεθειμένοι να αγοράσουν ένα προϊόν ή μια υπηρεσία.

Propensity based Segmentation

Βασίζεται σε τμηματοποιήσεις πελατών ανάλογα με τη βαθμολογία που έλαβαν (Scoring).

3.5 Market Basket Analysis

Η τεχνική Market Basket Analysis (MBA) είναι μια στατιστική μέθοδος, η οποία χρησιμοποιείται για να βοηθήσει τις επιχειρήσεις να αναλύσουν την συμπεριφορά των πελατών και τις καταναλωτικές τους κινήσεις. Ιστορικά οι κανόνες συσχέτισης αποτελούν βασική εργασία για την

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

διαδικασία εξόρυξης δεδομένων. Στο Market Basket Analysis δίνεται έμφαση στην ανάλυση του καλαθιού αγορών των πελατών στοχεύοντας στη συχνή εμφάνιση δεδομένων και τη δημιουργία συσχετίσεων μεταξύ τους. Οι κανόνες συσχέτισης έχουν ρίζες από την θεωρία των συνόλων, καθώς οι έννοιες υποσύνολα, τομές κτλ αποτελούν μέρος τους.

Η ανάλυση του καλαθιού αγορών μελετά τις αγορές που πραγματοποίησαν οι πελάτες στο ίδιο κατάστημα. Το σύνολο των προϊόντων που περιέχονται στο καλάθι οδηγεί σε αποτελέσματα ομάδων που πωλούνται πιο συχνά σε σχέση με άλλες ομάδες. Τα αποτελέσματα προσδίδουν την δυνατότητα στις ομάδες πωλήσεων να διαμορφώσουν την προώθηση των προϊόντων, όπως και να μεταβληθεί η διαρρύθμιση σε ένα φυσικό κατάστημα ή να προτείνονται συσχετιζόμενα προϊόντα σε ένα ηλεκτρονικό κατάστημα. Η τοποθέτηση προϊόντων σε κοντινές θέσεις έχει αποδειχθεί ότι αυξάνει τις πωλήσεις. Για παράδειγμα, όταν ένας πελάτης αγοράσει ένα μπουκάλι κρασί, είναι αρκετά πιθανό να αγοράσει ξηρούς καρπούς ή πατατάκια εφόσον βρίσκονται σε κοντινά ράφια. Σε αντίθετη περίπτωση, είναι αρκετά πιθανό ότι ο πελάτης δε θα θυμηθεί ότι το κρασί ταιριάζει με ξηρούς καρπούς με αποτέλεσμα να μην αγοράσει το επιπλέον προϊόν, κάτι το οποίο τελικά ίσως το αγοράσει σε διαφορετική χρονική στιγμή ή από άλλο κατάστημα. Ένα άλλο παράδειγμα είναι η διενέργεια προσφορών συγγενών προϊόντων που επίσης αποφέρει περισσότερες πωλήσεις.

Η μέθοδος MBA υπόσχεται σημαντικά αποτελέσματα και εντοπίζει συσχετίσεις που συνεισφέρουν στην ανάλυση της καταναλωτικής συμπεριφοράς. Επίσης, συμβάλει στην αύξηση των εσόδων και βελτιώνει τις παρακάτω περιοχές:

1. Cross-selling and up-selling
2. Sales support and direct marketing
3. Loyalty programmes
4. Product placement and store appearance
5. Supplier chain
6. Product bundling

3.5.1 Πλεονεκτήματα χρήσης Market Basket Analysis

Η ανάλυση του MBA στοχεύει στον εντοπισμό σχέσεων μεταξύ των αγορών που πραγματοποιούν οι πελάτες και την απεικόνιση αυτών των συσχετίσεων που βοηθούν στη λήψη αποφάσεων σε στρατηγικές κινήσεις για την αύξηση των πωλήσεων. Μέσα από την ανάλυση οι ερευνητές αποκτούν τη γνώση για τα προϊόντα που αγοράζονται πιο συχνά ή εντοπίζουν ποια είναι εκείνα τα προϊόντα που μπορούν να ενταχθούν σε μια προωθητική ενέργεια.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Αναλυτικότερα τα πλεονεκτήματα του Market Basket Analysis:

1. Βοηθά στην ανάπτυξη θεωριών αφού επιτρέπει τη διαχείριση τεράστιων όγκων δεδομένων με σκοπό την ανάλυσή τους που θα οδηγήσουν σε συμπεράσματα και ορισμούς. Για παράδειγμα ο Gary Russell (1999), Καθηγητής έρευνας στο Marketing, σημείωσε ότι η τεχνική του MBA συνέβαλε ώστε οι ερευνητές να κατασκευάσουν μοντέλα αποφάσεων που περιέχουν προϊόντα από πολλές κατηγορίες (Aguinis, Forcum, & Joo, 2013).

2. Το MBA έχει το πλεονέκτημα να χρησιμοποιεί δεδομένα τα οποία είναι αδόμητα. Δεδομένα που συγκεντρώνονται σε τακτική βάση αλλά έχουν σημαντικότητα, μπορούν να χρησιμοποιηθούν και να παράξουν αποτελέσματα.

3. Επίσης, παράγει αποτελέσματα διαφορετικών χρονικών περιόδων και δεν περιορίζεται σε πρόσφατες εγγραφές. Συνεπώς, είναι δυνατόν να δημιουργηθούν θεωρίες για παλαιότερες χρονικές στιγμές.

4. Επιπλέον, η MBA μπορεί να αναλύσει δεδομένα πολυδιάστατων επιπέδων. Στον τομέα του Marketing, εάν η οργανωτική μονάδα χρησιμοποιείται ως επίπεδο ανάλυσης, τότε οι αναλυτές μπορούν να αντλήσουν κανόνες συσχέτισης που συνδέουν το χαμηλότερο επίπεδο (π.χ. μεμονωμένα εστιατόρια Subway) με μονάδες υψηλότερου επιπέδου (π.χ. ολόκληρη αλυσίδα εστιατορίων Subway σε μία συγκεκριμένη πολιτεία) και με ακόμη υψηλότερα επίπεδα ανάλυσης (π.χ. όλα τα εστιατόρια του Subway στις Ηνωμένες Πολιτείες) (Aguinis, Forcum, & Joo, 2013).

5. Τέλος, η εφαρμογή του είναι φιλική προς τους ενδιαφερόμενους, καθώς εύκολα μπορούν να κατανοήσουν τα αποτελέσματα, αφού οι μετρήσεις κυμαίνονται σε μια κλίμακα επί τις 100.

3.5.2 Εφαρμογές

Η πρώτη επαφή με το μοντέλο οδηγεί τον αναγνώστη να σκέφτεται μόνο συναλλαγές σε ένα Super Market. Η πραγματικότητα είναι ότι η ανάλυση μπορεί να εφαρμοστεί σε δεκάδες άλλους τομείς. Για παράδειγμα ένα ηλεκτρονικό κατάστημα χρησιμοποιεί το MBA για να προτείνει στους χρήστες της ιστοσελίδας να αγοράσουν προϊόντα βάσει κανόνων συσχέτισης σύμφωνα με το προϊόν που έχουν προσθέσει ήδη στο καλάθι αγορών τους. Σχετικές εφαρμογές περιγράφονται παρακάτω:

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

1. **Λιανεμπόριο**: Κλασσική εφαρμογή σε επιχειρήσεις λιανικού εμπορίου με προσδιορισμό των προϊόντων που αγοράζονται συχνά σε συγκεκριμένη περίοδο ή με μια σειρά επιλογής. Αποδίδει αποτελέσματα σχετικά με την τοποθέτηση προϊόντων σε κοντινά σημεία εντός του καταστήματος και τη δημιουργία προωθητικών ενεργειών σε προϊόντα που ανήκουν στο ίδιο group (π.χ. υλικά για πρωινό, επίσημο ανδρικό ντύσιμο με σακάκι, παντελόνι, γραβάτα κτλ.).

2. **Τράπεζες**: Πρωτοποριακή εφαρμογή, καθώς χρησιμοποιείται σε συναλλαγές πιστωτικών καρτών, δημιουργώντας προγράμματα επιβράβευσης αγορών (βλ. Eurobank επιστροφή, Alpha bonus, κτλ). Επίσης συναντάται σε ενέργειες Cross-selling για την προώθηση επιπλέον προϊόντων. Για παράδειγμα, σε πελάτη με στεγαστικό δάνειο και ασφάλεια πυρός κατοικίας, η οποία είναι υποχρεωτική, προωθείται ασφάλεια κλοπής περιεχομένου και άλλων ειδών ασφάλειες.

3. **Ασφάλιση**: Εφαρμόζονται οι τεχνικές για τον εντοπισμών υποθέσεων με ενδεχόμενο απάτης (fraud detection).

4. **Τηλεπικοινωνίες**: Αναλύονται υπηρεσίες που χρησιμοποιούν οι πελάτες δίνοντας την δυνατότητα για νέες προωθητικές ενέργειες που αποδίδουν κίνητρο και αποτρέπουν τους συνδρομητές να αλλάξουν πάροχο.

5. **Ιατρική**: η ανάλυση συμπτωμάτων δημιουργεί ιατρικά profiles τα οποία διευκολύνουν την εφαρμογή επιτυχημένων θεραπειών.

3.5.3 Market Basket Analysis ορισμοί

Γενικά η συμπεριφορά του καταναλωτή είναι μια εγκεφαλική δραστηριότητα που σχετίζεται με τις αποφάσεις να αγοράσει, να χρησιμοποιήσει, να διερευνήσει την αγορά. Επειδή κάθε πελάτης έχει διαφορετικές προτιμήσεις και ανάγκες, αυτό δεν σημαίνει ότι εντοπίζονται ομοιότητες. (Kurniawana, Umayaha, Hammad, Susiki Nugroho, & Hariadi, 2018). Αυτές οι συμπεριφορές μπορούν να προσδιοριστούν μέσα από την ανάλυση αρχείων συναλλαγών εμπορικών οργανισμών. Για τον παραπάνω λόγο χρησιμοποιείται κατά κόρον η τεχνική του Market Basket Analysis.

Τα Association Rules σχετίζονται με ερωτήματα τι, ποιο, με ποιο άλλο. Καθώς η εφαρμογή τους στα Data Analytics έχει άμεση σχέση με την ανάλυση συναλλαγών μιας επιχείρησης, πολύ συχνά τα Association rules αναφέρονται ως Market Basket Analysis.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Αναλυτικότερα το MBA μπορεί να απαντήσει σε ερωτήσεις τύπου:

1. Που θα τοποθετηθούν τα προϊόντα στα ράφια;
2. Ποιοι συνδυασμοί προϊόντων θα συμπεριληφθούν σε προωθητικές ενέργειες;
3. Προσδιορίζει την τιμολογιακή πολιτική προϊόντων ίδιας κατηγορίας.

Οι κανόνες συσχέτισης που παράγονται από την εφαρμογή του μοντέλου διαφέρουν ως προς την σημαντικότητά τους. Για τον προσδιορισμό της σημαντικότητας λαμβάνονται υπόψη αρκετές παράμετροι. Οι δυο σημαντικότερες παράμετροι που παρουσιάζονται στην έρευνα είναι το support και το confidence.

Support

Το Support περιγράφει την πιθανότητα τα προϊόντα Α και Β να βρίσκονται στο ίδιο καλάθι αγορών.

$$Support = P\ (A\ \cap B) = \frac{n(A \cup B)}{N}$$

$$= \frac{(Αριθμός\ συναλλαγών\ που\ περέχουν\ τα\ προιόντα\ A\ και\ B)}{(Συνολικός\ αριθμός\ συναλλαγών)}$$

Confidence

Πρόκειται για την τιμή που προσδιορίζει την ισχύ που έχουν μεταξύ τους τα προϊόντα. Στατιστικά είναι η δεσμευμένη πιθανότητα να αγοράσει κάποιος το προϊόν Β δεδομένου ότι έχει αγοράσει ήδη το προϊόν Α.

$$Confidence = P(A|B) = \frac{Αριθμός\ συναλλαγών\ που\ περιέχουν\ τα\ προϊόντα\ A\ και\ B}{Σύνολο\ συναλλαγών\ που\ περιέχουν\ το\ προϊόν\ A}$$

Ωστόσο υπάρχουν κι άλλες παράμετροι που μπορούν να συνεκτιμηθούν:

Expected Confidence

Περιγράφει την πιθανότητα του προϊόν Β δεδομένου του προϊόντος Α, όταν αυτά τα δυο ενδεχόμενα είναι μεταξύ τους ανεξάρτητα.

$$Expected\ Confidence = P(B|A) = \frac{n(A \cap B)}{n(A)}$$

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Lift

Το Lift είναι μια τιμή πολύ σημαντική για τους κανόνες και είναι η ένταση μεταξύ των σχέσεων που δημιουργεί το λογισμικό. Πιο συγκεκριμένα το Lift ενός κανόνα συσχέτισης είναι ο λόγος Confidence προς το Expected Confidence.

Επειδή το Lift είναι ένας λόγος, η τιμή που μπορεί να λάβει ενδέχεται να είναι μεγαλύτερη ή μικρότερη του 1.

$$Lift = \frac{Support\ (A\ \&\ B)}{Support(A) * Support(B)}$$

Συνεπώς σαν αποτέλεσμα:

- Εάν το Lift έχει τιμή μεγαλύτερη από 1, σημαίνει ότι η απόκριση στόχος είναι πιο πιθανή από τη μέση απόκριση. Επομένως, ο κανόνας συσχέτισης βελτιώνει τις πιθανότητες του αποτελέσματος.
- Εάν είναι μικρότερο από 1, τότε η απόκριση στόχος είναι λιγότερο πιθανή από τη μέση απόκριση. Ως αποτέλεσμα, ο κανόνας συσχέτισης μειώνει τις πιθανότητες για το επιθυμητό αποτέλεσμα.
- Εάν το Lift ισούται με 1, σημαίνει ότι το μοντέλο (ή ο κανόνας συσχέτισης) δεν επηρεάζει το αποτέλεσμα.

Για παράδειγμα, εάν το Lift είναι 5.6, τότε ο πελάτης που έχει αγοράσει το προϊόν Α, είναι 5.6% πιθανότερο να αγοράσει και το Β προϊόν. Κατά την εκτέλεση της ανάλυσης το λογισμικό θα παράξει αρκετούς κανόνες σε πολλές συναλλαγές. Για αυτό τον λόγο θέτουμε thresholds στο Confidence και στο Support. Για παράδειγμα, τα Association Rules που θα δημιουργηθούν μόνο όταν το Support είναι μεγαλύτερο του 5% και το Confidence μεγαλύτερο του 10%, ώστε να μην δημιουργηθεί σύγχυση στους λήπτες αποφάσεων. Στην περίπτωση που εξάγονται αρκετοί κανόνες, ο ερευνητής μπορεί σταδιακά να αυξήσει το Confidence και το Support, έτσι ώστε να μειωθούν τα αποτελέσματα του λογισμικού.

3.5.4 Consumer Behaviour

Η συμπεριφορά των καταναλωτών ορίζεται ως η μελέτη ατόμων, ομάδων ή οργανισμών και των διαδικασιών που χρησιμοποιούν για την επιλογή, τη χρήση και την απόρριψη προϊόντων, υπηρεσιών, εμπειριών ή ιδεών για την κάλυψη των αναγκών και του αντίκτυπου αυτής της διαδικασίας στους καταναλωτές και την κοινωνία. Η μελέτη αναφέρεται στις ενέργειες των

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

καταναλωτών στην αγορά και στα κίνητρα που διέπουν αυτές τις ενέργειες (Alfarizi & Sari, 2021). Η συμπεριφορά των καταναλωτών επηρεάζεται από διάφορους παράγοντες όπως η κουλτούρα, κοινωνικούς, ψυχολογικούς και προσωπικούς παράγοντες, οι οποίοι συνοψίζονται στον παρακάτω πίνακα:

Παράγοντες	**Περιγραφή**
Κουλτούρα	Σε προϊόντα που αφορούν την οικογένεια, την καταγωγή, την εθνικότητα, την γεωγραφική θέση, κτλ.
Κοινωνικοί	Κοινωνική θέση, ομάδες ατόμων, οικογένεια, φίλοι, επαγγελματική αναγνώριση
Προσωπικοί	Οικογενειακή κατάσταση, ηλικία, επαγγελματικό επίπεδο, οικονομική κατάσταση
Ψυχολογικοί	Αίσθηση ότι με την απόκτηση αγαθών ικανοποιείται κάποια ψυχολογική ανάγκη, ανάγκη για προμήθεια αγαθών σε περίπτωση έκτακτης ανάγκης

Πίνακας 3 - Customer Behavior Factors

3.5.5 General Data Protection Regulation (GDPR)

Τον Μάιο του 2016 δημοσιεύθηκε ο Γενικός Κανονισμός Προστασίας Δεδομένων της Ε.Ε., ευρύτερα γνωστός ως GDPR (General Data Protection Regulation - 2016/679). Σκοπός της δημοσίευσής του υπήρξε η θέσπιση ομοιόμορφου νομικού πλαισίου από τα κράτη-μέλη της Ε.Ε. για την προστασία των φυσικών προσώπων έναντι της επεξεργασίας δεδομένων προσωπικού χαρακτήρα. Ο Κανονισμός επέφερε ίσως τη μεγαλύτερη αλλαγή στην νομοθεσία για τη διαφύλαξη και προστασία των υποκειμένων των δεδομένων (data subjects) έναντι της επεξεργασίας των προσωπικών τους δεδομένων τη τελευταία εικοσαετία εντός της Ε.Ε.

Συγκεκριμένα, ο Κανονισμός έχει πεδίο εφαρμογής σε επιχειρήσεις με έδρα εντός ή/και εκτός Ε.Ε. με κριτήριο τον τόπο εγκατάστασης της επιχείρησης και επεξεργασίας των προσωπικών δεδομένων ή/και τον τόπο διαμονής των υποκειμένων των δεδομένων. Επιπλέον, αφορά φορείς τόσο του ιδιωτικού, όσο και του δημόσιου τομέα (ιδιωτικές και δημόσιες επιχειρήσεις, συλλόγους, κρατικές αρχές, κλπ.) οι οποίοι διαχειρίζονται, διακινούν, αποθηκεύουν και εν γένει επεξεργάζονται δεδομένα προσωπικού χαρακτήρα φυσικών προσώπων, τα οποία αφορούν Ευρωπαίους πολίτες ή σχετίζονται με υπηρεσίες και αγαθά προσφερόμενα προς Ευρωπαίους πολίτες.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Ο Γενικός Κανονισμός Προστασίας Δεδομένων αν και έχει σαφώς μεγαλύτερη κρισιμότητα για μεγάλες επιχειρήσεις (εταιρείες που απασχολούν περισσότερους από 250 εργαζομένους), δεν εξαιρεί τις μικρομεσαίες επιχειρήσεις (SMEs). Επομένως, ακόμα και οι τελευταίες που διακινούν, αποθηκεύουν και εν γένει επεξεργάζονται προσωπικά δεδομένα, τα οποία είναι είτε σε φυσική μορφή (έντυπα, χειρόγραφα αρχεία), είτε σε ψηφιακή (στοιχεία συνεργατών ή πελατών e-shop, marketing web sites, συστήματα CRM) πρέπει να εφαρμόζουν απαρέγκλιτα τις διατάξεις του Κανονισμού.

Έτσι, λοιπόν, οι ιδιοκτήτες καταστημάτων ηλεκτρονικού εμπορίου πρέπει μεταξύ των άλλων θεμάτων που διαχειρίζονται (digital marketing, social media, web design, κλπ.), να κινούν και τις απαραίτητες ενέργειες για τη συμμόρφωση της επιχείρησής τους με τις διατάξεις του Κανονισμού.

Τα προσωπικά δεδομένα που μπορεί να συλλέγει και να επεξεργάζεται ένα e-shop σχετίζονται με οποιαδήποτε πληροφορία που αφορά ένα ταυτοποιημένο ή ταυτοποιήσιμο φυσικό πρόσωπο, είτε αυτό κατοικεί σε κράτος μέλος της Ε.Ε., είτε συμμετέχει σε κάποια επιχείρηση της Ε.Ε. Έτσι, οποιαδήποτε πληροφορία μπορεί να χρησιμοποιηθεί για την άμεση ή έμμεση ταυτοποίηση ενός φυσικού προσώπου, ιδίως μέσω αναφοράς σε αναγνωριστικό στοιχείο ταυτότητας, σχετίζεται με τα προσωπικά δεδομένα και την επεξεργασία αυτών.

Στην πράξη, όταν ένας πελάτης καταχωρεί το ονοματεπώνυμό του ή τη διεύθυνση ηλεκτρονικού ταχυδρομείου στην πλατφόρμα κάποιου e-shop, το εν λόγω e-shop συλλέγει και αποθηκεύει προσωπικά δεδομένα. Το ίδιο ισχύει και όταν ο ιδιοκτήτης/υπάλληλος/υπεύθυνος αναλυτής του e-shop παρακολουθεί (επεξεργάζεται) τις διευθύνσεις IP για να ταξινομήσει τους επισκέπτες ανά χώρα, ηλικία, κλπ. Περιληπτικά, οποιαδήποτε πράξη ή σειρά πράξεων που πραγματοποιείται με ή χωρίς τη χρήση αυτοματοποιημένων μέσων σε προσωπικά δεδομένα ή σε σύνολα προσωπικών δεδομένων, όπως η συλλογή, η καταχώριση, η αποθήκευση, η προσαρμογή ή η μεταβολή θεωρείται επεξεργασία προσωπικών δεδομένων. Χαρακτηριστικά παραδείγματα προσωπικών δεδομένων είναι:

- Όνομα, επώνυμο
- Διεύθυνση
- Τηλεφωνικός αριθμός, FAX, skype, viber, κλπ
- Φωτογραφία, video
- Ημερομηνία γέννησης
- Αριθμός ταυτότητας, ΑΦΜ, κ.λπ.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

- E-mail
- Online αναγνωριστικά (ψηφιακά αναγνωριστικά όπως η διεύθυνση IP, τα cookies, το MAC address ή το IMEI ενός κινητού)

Τα e-shops όπως και όλες οι επιχειρήσεις σύμφωνα με τα παραπάνω που έχουν την ιδιότητα του Υπευθύνου Επεξεργασίας ή του Εκτελούντος την Επεξεργασία σύμφωνα με τις διατάξεις του Κανονισμού θα πρέπει να λαμβάνουν όλα εκείνα τα απαραίτητα τεχνικά και οργανωτικά μέτρα, προκειμένου να διασφαλίζουν επαρκές επίπεδο προστασίας των προσωπικών δεδομένων που επεξεργάζονται στο πλαίσιο των δραστηριοτήτων τους, την εμπιστευτικότητα, την ακεραιότητα και τη διαθεσιμότητα αυτών (Το παραπάνω κείμενο επιβλέφθηκε από τον Αλεβιζόπουλο Χαράλαμπο Data Protection Officer της Τράπεζας ΟΡΤΙΜΑ).

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

4

Πώς οι τεχνικές data analytics συμβάλουν στην αύξηση των εσόδων των ηλεκτρονικών καταστημάτων

Στα πρώτα χρόνια ύπαρξης των διαδικτυακών ιστοσελίδων το ζήτημα ήταν μια επιχείρηση να καταφέρει να προβληθεί σε υποψήφιους πελάτες, οι οποίοι δεν βρίσκονται κοντά στο φυσικό κατάστημα. Με τη νέα τεχνολογία για την εποχή δινόταν η δυνατότητα να γίνει ευρύτερα γνωστή η επιχείρηση και με αυτόν τον τρόπο να αυξήσει τα έσοδα της και, κατ΄ επέκταση, τα κέρδη της. Με το πέρασμα των δεκαετιών και την ανάπτυξη της τεχνολογίας, οι ανάγκες διαφοροποιήθηκαν. Η ραγδαία εξέλιξη της τεχνολογίας, ιδιαίτερα των συστημάτων διαχείρισης databases και data mining, των δικτύων και η επεξεργασία των πληροφοριών έφεραν αλλαγές στις αγορές και στο Marketing. Τα ηλεκτρονικά καταστήματα λειτουργούν γρηγορότερα από το παρελθόν με ανεπτυγμένη οργάνωσή σε όλα τα επίπεδα, όπως το Marketing, τη λήψη της παραγγελίας μέχρι και την αποστολή της. Όμως ακόμα περισσότερο, η πρόσβαση σε τεράστιο όγκο πληροφοριών από τις ενέργειες των πελατών με ευκολία υπόσχεται ακόμα περισσότερα συμπεράσματα, τα οποία οδηγούν σε νέες συνήθως καινοτόμες λύσεις. Ο ανταγωνισμός των επιχειρήσεων είναι πολύ μεγάλος και προκειμένου να εδραιωθούν στην αγορά αναζητούν συνεχώς τις επόμενες κινήσεις τους. Πλέον σημασία έχουν μετρήσιμα στοιχεία, όπως είναι ο αριθμός επισκέψεων από τους πελάτες, ποια προϊόντα είδαν, κ.α. Κατά τη χρήση μιας ιστοσελίδας αποθηκεύεται ένας μεγάλος όγκος πληροφοριών που μπορεί να οδηγήσει στην παραγωγή πλήθους αποτελεσμάτων και συμπερασμάτων. Σήμερα υπάρχει η δυνατότητα να γνωρίζει η επιχείρηση:

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

- Το μέσο χρόνο παραμονής ενός πελάτη
- Ποια προϊόντα είναι πιο δημοφιλή
- Από ποιόν σύνδεσμο πραγματοποιήθηκε η σύνδεση (π.χ. Google Search, Advertisement, etc)
- Τον αριθμό των επισκέψεων
- Τον αριθμό των παραγγελιών

Από όλα τα παραπάνω μπορούν να παραχθούν διάφορα συμπεράσματα για την επιτυχία ενός ηλεκτρονικού καταστήματος. Ωστόσο δεν αρκεί μόνο η μέτρηση των καταγραφών. Οι ανάγκες των επιχειρήσεων είναι πολλές περισσότερες και προκειμένου να αυξήσουν ή να διατηρήσουν το πελατολόγιό τους, χρειάζονται λύσεις εξατομικευμένων προωθητικών ενεργειών προς τους πελάτες, οι οποίες θα αποτελέσουν μια προσωπική εμπειρία για τον καθένα από αυτούς.

Οι τεχνικές data mining μπορούν να απαντήσουν ερωτήματα όπως:

1. Ποιοι πελάτες έχουν τη μεγαλύτερη/λιγότερη αξία για μια επιχείρηση;
2. Ποια είναι τα ιδιαίτερα χαρακτηριστικά τους;
3. Ποιοι είναι οι περισσότερο/λιγότερο πιστοί πελάτες και πώς χαρακτηρίζονται;
4. Ποια προϊόντα/αντικείμενα αγοράζουν συχνά οι πελάτες μαζί;
5. Με ποια σειρά έχουν αγοραστεί τα προϊόντα;
6. Ποια κατηγορία πελατών είναι πιθανότερο να ανταποκριθούν σε ένα promotion;

Οι ερευνητές στο χώρο του Marketing αναζητούν με την εφαρμογή νέων τεχνολογιών την επίλυση προβλημάτων που αφορούν τον τομέα όπως το Segmentation, το Prediction και τα συστήματα υποστήριξης αποφάσεων (Hiziroglu, 2013).

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

5

Η Επιχείρηση

5.1 Business Background

Η επιχείρηση που εξετάζουμε στην παρούσα διπλωματική λειτουργεί ως ηλεκτρονικό κατάστημα από το 2016 και κατηγοριοποιείται στις μικρομεσαίες επιχειρήσεις στην Ελλάδα. Διατηρεί ωστόσο και φυσικό κατάστημα στην Αθήνα με αποθήκες κι έχει παρουσία στο χώρο περισσότερο από 20 χρόνια. Η εν λόγω επιχείρηση δραστηριοποιείται στον κλάδο των εποχιακών ειδών και εμπορεύεται τουλάχιστον 5.000 κωδικούς προϊόντων. Απασχολεί 10-12 υπαλλήλους σε εργασίες λογιστηρίου, εξυπηρέτησης πελατών και λήψης και οργάνωσης παραγγελιών προς αποστολή. Υπενθυμίζεται ότι για λόγους προστασίας τόσο της επιχείρησης, όσο και των προσωπικών δεδομένων των πελατών της [βλέπε ενότητα 3.5.5 General Data Protection Regulation (GDPR)], δεν αναφέρεται πουθενά στη διπλωματική η επωνυμία της επιχείρησης και πιθανά άλλα στοιχεία τα οποία μπορεί να αποκαλύψουν την ταυτότητά της.

Οι κύριες κατηγορίες των προϊόντων που εμπορεύεται η επιχείρηση είναι:

- Χριστουγεννιάτικα είδη
- Αποκριάτικα είδη
- Είδη κήπου και βεράντας
- Είδη σπιτιού

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

5.2 Παρουσίαση Dataset

Για τις ανάγκες της εργασίας η συνεργαζόμενη εταιρία προσέφερε τέσσερα αρχεία δεδομένων με αριθμητικά στοιχεία πελατών, παραγγελίες και τιμολογημένες παραγγελίες με ιδιαίτερη προσοχή, όπως προαναφέραμε για την αποφυγή παραβίασης των προσωπικών δεδομένων του πελατολογίου της.

5.2.1 Data Set Customers

Το σύνολο των δεδομένων αφορούν πελάτες, οι οποίοι μέσω της Ιστοσελίδας του e-shop πραγματοποίησαν εγγραφή ως χρήστες και στη συνέχεια προχώρησαν σε κάποια παραγγελία/διαδικτυακή αγορά. Συγκεκριμένα, το σύνολο των πελατών που περιέχονται στο εν λόγω αρχείο είναι 7.064.

Το data set έχει τα παρακάτω χαρακτηριστικά:

- virtuemart_order_userinfo_id: Κωδικός καταχώρησης στοιχείων πελάτη στη βάση
- virtuemart_order_id: Κωδικός παραγγελίας (κλειδί για σύνδεση με το αρχείο παραγγελιών)
- Customer_ID: Μοναδικοποιημένος κωδικός πελάτη στη βάση δεδομένων
- city: Πόλη
- zip: Ταχυδρομικός Κώδικας
- Gender: Φύλο πελάτη (για Νομικά πρόσωπα η τιμή είναι «Ο»)
- Date_of_Birth: Καταχωρημένη ημερομηνία γέννησης από τον πελάτη

Inc	Source Name	Name	Label	Type	Source Informat	Len.
☑	virtuemart_order_userinfo_id	virtuemart_order_userinfo_id	virtuemart_order_userinfo_id	String	$CHAR6.	6
☑	virtuemart_order_id	virtuemart_order_id	virtuemart_order_id	Number	BEST4.	8
☑	Customer_ID	Customer_ID	Customer_ID	Number	BEST4.	8
☑	city	city	city	String	$CHAR122.	122
☑	zip	zip	zip	String	$CHAR58.	58
☑	Gender	Gender	Gender	String	$CHAR1.	1
☑	Date_of_Birth	Date_of_Birth	Date_of_Birth	String	$CHAR10.	10

Πίνακας 4 - Dataset "Customers"

5.2.2 Data Set Orders Items

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Σε αυτό το αρχείο περιέχονται οι συναλλαγές με τα προϊόντα από κάθε παραγγελία. Οι συνολικές συναλλαγές είναι 8.399. Το αρχείο περιέχει μια εγγραφή για κάθε SKU, η οποία αντιστοιχεί σε κάθε απόδειξη.

Συγκεκριμένα, περιέχει τις παρακάτω κατηγορίες:

- virtuemart_order_item_id: Κωδικός καταχώρησης στη βάση δεδομένων
- virtuemart_order_id: Κωδικός παραγγελίας (κλειδί για σύνδεση με το αρχείο παραγγελιών)
- order_item_sku: Κωδικός προϊόντος
- order_item_name: Όνομα προϊόντος
- product_quantity: Ποσότητα
- product_final_price: Αξία προϊόντος
- product_subtotal_with_tax: Συνολικό Κόστος (Ποσότητα * Αξία προϊόντος)
- Category: Κατηγορία στην οποία ανήκει το προϊόν

Inc	Source Name	Name	Label	Type	Source Informat	Len.
☑	virtuemart_order_item_id	virtuemart_order_item_id	virtuemart_order_item_id	Number	BEST5.	8
☑	virtuemart_order_id	virtuemart_order_id	virtuemart_order_id	Number	BEST4.	8
☑	order_item_sku	order_item_sku	order_item_sku	String	$CHAR78.	78
☑	order_item_name	order_item_name	order_item_name	String	$CHAR303.	303
☑	product_quantity	product_quantity	product_quantity	Number	BEST2.	8
☑	product_final_price	product_final_price	product_final_price	Number	COMMA10.	8
☑	product_subtotal_with_...	product_subtotal_with_tax	product_subtotal_with_tax	Number	COMMA10.	8

Πίνακας 5 - Dataset "Orders Items"

5.2.3 Data Set Orders payments analysis

Ο συγκεκριμένος πίνακας περιέχει την οικονομική ανάλυση του κόστους κάθε παραγγελίας και περιέχει τις κάτωθι κατηγορίες:

- virtuemart_order_id: Κωδικός καταχώρησης παραγγελίας στη βάση δεδομένων
- order_number: Κωδικός συστήματος
- order_total: Τελικό ποσό παραγγελίας
- order_salesPrice: Τελικό ποσό εμπορευμάτων
- order_subtotal: Τελικό ποσό εμπορευμάτων χωρίς ΦΠΑ
- order_tax: ΦΠΑ
- order_shipment: Χρέωση μεταφορικών
- order_payment: Έξτρα χρέωση για πληρωμή (π.χ. αντικαταβολή)

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

- virtuemart_paymentmethod_id: Τρόπος πληρωμής
- virtuemart_shipmentmethod_id: Αποστολή
- created_on: Ημ/νια Παραγγελίας

Inc	Source Name	Name	Label	Type	Source Informat	Len.
☑	virtuemart_order_id	virtuemart_order_id	virtuemart_order_id	Number	BEST4.	8
☑	order_number	order_number	order_number	String	$CHAR9.	9
☑	order_total	order_total	order_total	Number	COMMA10.	8
☑	order_salesPrice	order_salesPrice	order_salesPrice	Number	COMMA10.	8
☑	order_subtotal	order_subtotal	order_subtotal	Number	COMMA10.	8
☑	order_tax	order_tax	order_tax	Number	COMMA9.	8
☑	order_shipment	order_shipment	order_shipment	Number	COMMA8.	8
☑	order_payment	order_payment	order_payment	Number	COMMA4.	8
☑	virtuemart_payme...	virtuemart_payme...	virtuemart_paymentmethod...	Number	BEST1.	8
☑	virtuemart_shipme...	virtuemart_shipme...	virtuemart_shipmentmethod...	Number	BEST2.	8
☑	created_on	created_on	created_on	Date/Time	ANYDTDTM16.	8

Πίνακας 6 - Dataset "Orders Payment Analysis"

5.2.4 Data Set Payment method

Το αρχείο αυτό περιέχει τους τύπους πληρωμών των παραγγελιών και αντιστοιχεί στο Data Set "Orders payment analysis" με την κατηγορία «virtuemart_paymentmethod_id» και περιέχει τις εξής κατηγορίες:

- Κωδικός_Πληρωμής: ο κωδικός πληρωμής (κλειδί για με το αρχείο payment analysis)
- Περιγραφή τρόπου πληρωμής: περιγραφή του τρόπου πληρωμής (πχ τραπεζική κατάθεση κτλ)

Inc	Source Name	Name	Label	Type	Source Informat	Len.
☑	Κωδικός_πληρωμής	Κωδικός_πληρωμής	Κωδικός_πληρωμής	Number	BEST1.	8
☑	Περιγραφή_τρόπου_πληρ...	Περιγραφή_τρόπου_	Περιγραφή_τρόπου_πληρωμής:	String	$CHAR79.	79

Πίνακας 7 - Data Set Payment method

5.3 Ανάλυση Πελατολογίου

Όπως αναλύσαμε νωρίτερα, η επιχείρηση λαμβάνει καλύτερες αποφάσεις όταν γνωρίζει το πελατολόγιο στο οποίο απευθύνεται. Ξεκινώντας, λοιπόν, το τεχνικό κομμάτι της διπλωματικής,

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

χρησιμοποιήσαμε το πρόγραμμα SAS Enterprise Guide για την επεξεργασία και την ανάλυση των δεδομένων. Ακολουθούν κάποια πρώτα συμπεράσματα αναφορικά με το πελατολόγιο της επιχείρησης.

Αρχικά, από το παρακάτω Bar Chart αντιλαμβανόμαστε ότι το μεγαλύτερο ποσοστό των πελατών είναι γυναίκες, πλησιάζοντας σε ποσοστό το 63%, λίγο περισσότερο από το 36% είναι άνδρες και μόλις το 1% αφορά πελάτες επιχειρήσεις Νομικής μορφής.

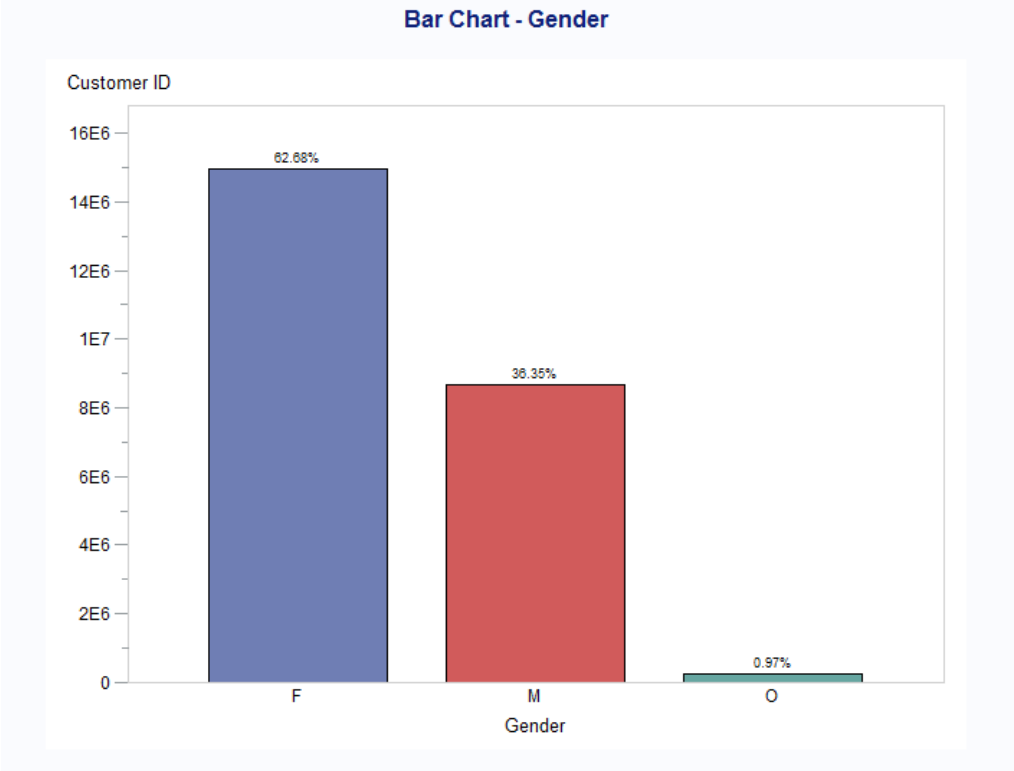


Chart 4 - Customers per Gender

Label	Value	Frequency Count	Percent of Total Frequency
Gender	F	4680	66.2514
	M	2299	32.5453
	O	85	1.2033

Πίνακας 8 - Gender Frequency Report Table

Από την πληροφορία της ηλικίας ο ερευνητής συγκεντρώνοντας σε ομάδες ηλικιών τους πελάτες, λαμβάνει περισσότερη πληροφόρηση για τα groups των ηλικιών που επιλέγουν την επιχείρηση για τις αγορές τους.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Διαπιστώνεται ότι το μεγαλύτερο ποσοστό των πελάτων ανήκουν στο ηλικιακό group 30-40 με ποσοστό 35%, ενώ το αμέσως επόμενο group με ποσοστό σχεδόν 25% είναι εκείνο των 50-60. Ακολουθεί το Group 40-50 με 23% και τέλος, το Group 60-70 με 11%. Αίσθηση προκαλεί το group 18-30, το οποίο συγχωνεύεται με το group 70+, καθώς συγκεντρώνει 6% ποσοστό. Από την πληροφορία αυτή συμπεραίνουμε ότι οι ηλικίες μικρότερες των 30 ετών δεν έχουν την τάση να αναζητούν προϊόντα διακοσμητικά, χριστουγεννιάτικα διακοσμητικά, έπιπλα κήπου κτλ. Το γεγονός αυτό πιθανόν να οφείλεται στην περιορισμένη οικονομική δυνατότητα λόγω ανεργίας και χαμηλών μισθών στο συγκεκριμένο ηλικιακό group ή ακόμα και στο γεγονός ότι απευθύνονται σε πιο δημοφιλή καταστήματα.

Όσον αφορά στη γεωγραφική ανάλυση και κατηγοριοποίηση του πελατολογίου έχουμε τα εξής αποτελέσματα:

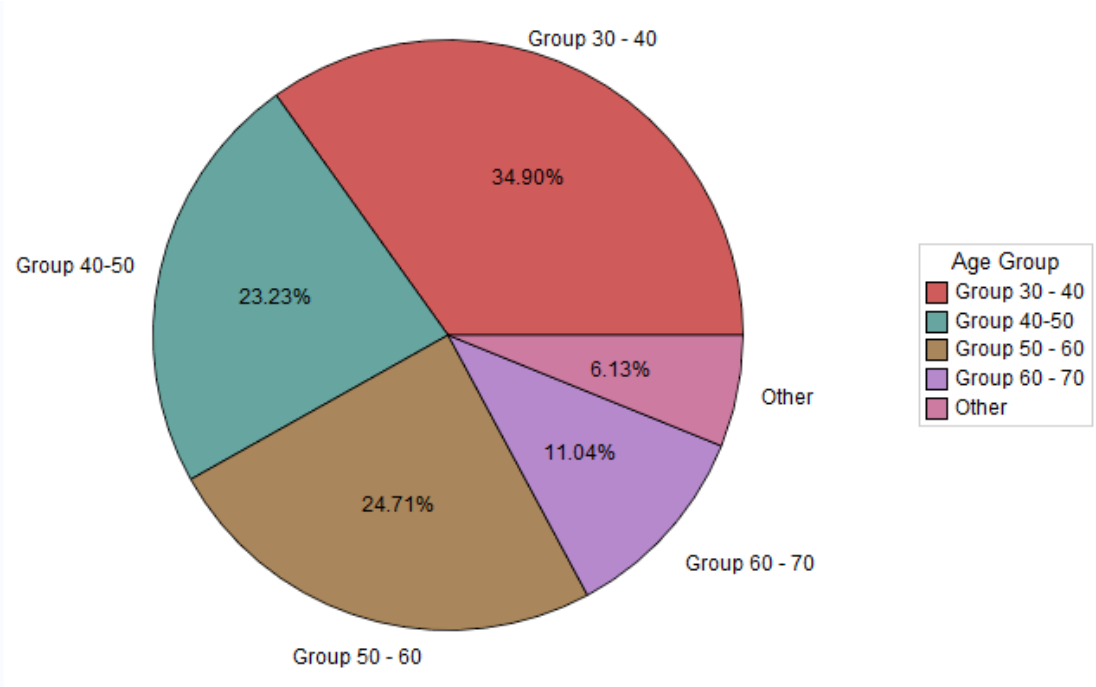


Chart 5 - Customers Group of Ages

Label	Value	Frequency Count	Percent of Total Frequency
Age Group	Group 30 - 40	2.739	38,77
	Group 40-50	1.320	18,68
	Group 18-30	1.252	17,72
	Group 50 - 60	1.161	16,43

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Label	Value	Frequency Count	Percent of Total Frequency
	Group 60 - 70	493	6,97
	Under 18	78	1,10
	Group 70+	21	0,29

Πίνακας 9 - Group of Ages Frequency Report Table

Η επιχείρηση, όπως προαναφέραμε, έχει έδρα σε περιοχή της Αθήνας και από το παρακάτω donut chart προκύπτει ότι το 40% των πελατών που παρήγγειλαν έστω και μια φορά από το e-shop, έχουν έδρα στην Αττική. Οι περιφέρειες της Θεσσαλίας, της Ανατολικής και Δυτικής Μακεδονίας, της Ηπείρου, των Ιονίων Νήσων και των Νήσων του Βορείου Αιγαίου συγκεντρώνουν μικρά ποσοστά και για το λόγο αυτό απεικονίζονται σε κομμάτι με ποσοστό μεγαλύτερο από 16%. Ακολουθούν η Κεντρική Μακεδονία με σχεδόν 13% και στη συνέχεια οι περιφέρειες Νοτίου Αιγίου, Στερεάς Ελλάδας, Κρήτης, Δυτικής Ελλάδας και Πελοποννήσου. Διαπιστώνεται ότι η εμπορευσιμότητα και αναγνωρισιμότητα της επιχείρησης δεν περιορίζεται στην Αττική που είναι και το φυσικό κατάστημα, αλλά έχει πανελλαδική εμβέλεια και αυτό επιτυγχάνεται με την λειτουργία του διαδικτυακού ηλεκτρονικού καταστήματος.

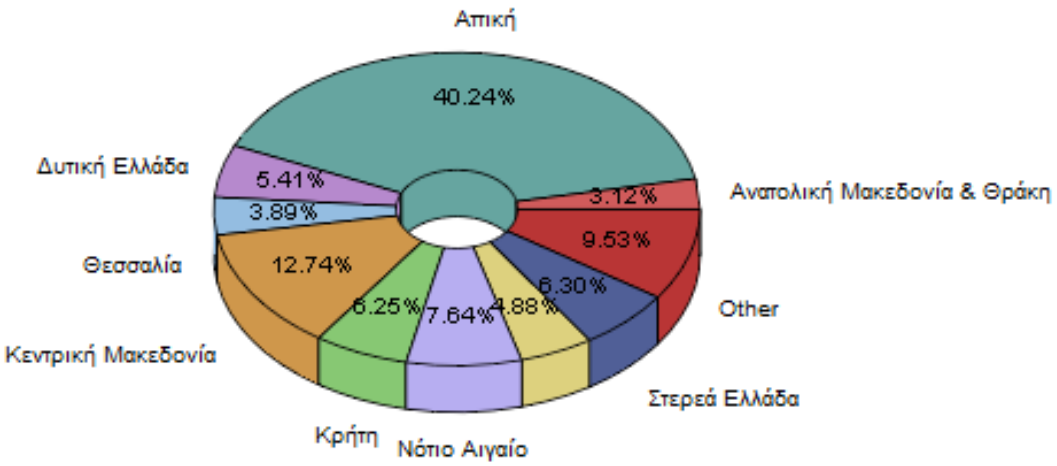


Chart 6 - Πελάτες ανά Διοικητική Περιφέρεια

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

5.4 Πωλήσεις

Όπως αναφέραμε και πιο πάνω, η πρωτοφανής κατάσταση με την πανδημία που δημιουργήθηκε από τις αρχές του 2020, δημιούργησε μια πληθώρα προβλημάτων σε επιχειρήσεις και λιανέμπορους. Το καταναλωτικό κοινό μετακινήθηκε σε αγορές μέσω διαδικτύου, αφού δεν ήταν εφικτή η παρουσία σε φυσικό κατάστημα. Τα παραπάνω αποτυπώνονται και στο Bar Chart που ακολουθεί. Συγκεκριμένα, από την έναρξη του e-shop το 2015, οι πωλήσεις ήταν ελάχιστες, πλησιάζοντας σχεδόν τα 2.000€ σε τζίρο. Τα έτη 2016 έως 2019 διαπιστώνεται ότι υπάρχει μια σταδιακά αυξητική τάση. Όμως, το 2020 και με την έναρξη της πανδημίας οι πωλήσεις εκτοξεύονται, σημειώνοντας αύξηση 490%. Το 2021 συνεχίζει η αυξητική πορεία, με μικρότερο, όπως είναι λογικό ρυθμό.

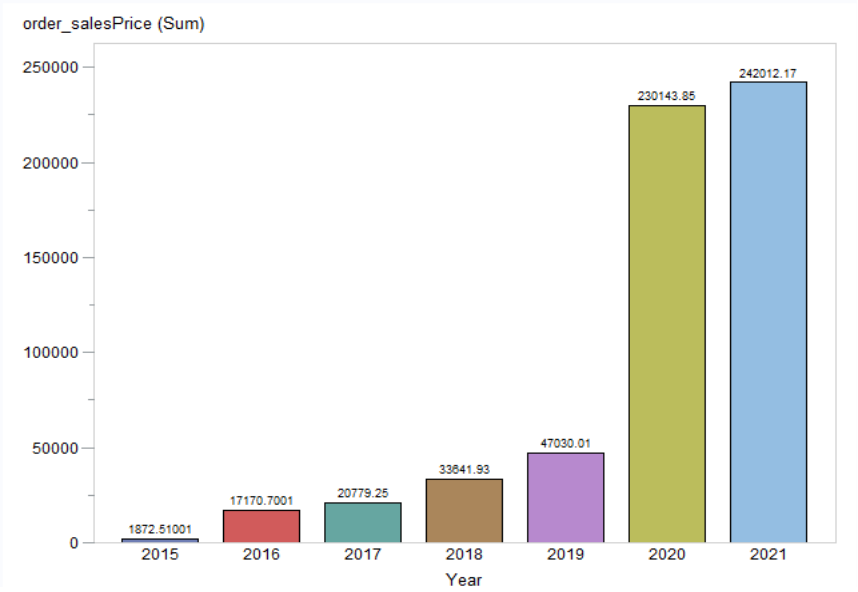


Chart 7 - Πωλήσεις ανά έτος

Η εταιρία, ως σύγχρονο ηλεκτρονικό κατάστημα, δίνει στους πελάτες της τη δυνατότητα διαφορετικών τρόπων πληρωμής. Ο πελάτης μπορεί να καταθέσει χρήματα σε τραπεζικούς λογαριασμούς στις τέσσερις μεγάλες τράπεζες της Ελλάδας, να πληρώσει με αντικαταβολή, αλλά και με πιστωτική κάρτα και τέλος, μέσω χρηματοπιστωτικών υπηρεσιών και συγκεκριμένα, τη χρήση ηλεκτρονικού πορτοφολιού από γνωστό πάροχο στην ελληνική αγορά.

Από το παρακάτω chart αντιλαμβάνεται ο ερευνητής ότι οι πελάτες προτιμούν να εξοφλούν τις παραγγελίες τους μέσω Τραπεζικής κατάθεσης σε ποσοστό 37%, με σχεδόν 36% εξοφλούν με πιστωτική κάρτα, ενώ σημαντικά μεγάλο ποσοστό πελάτων πληρώνει με αντικαταβολή.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

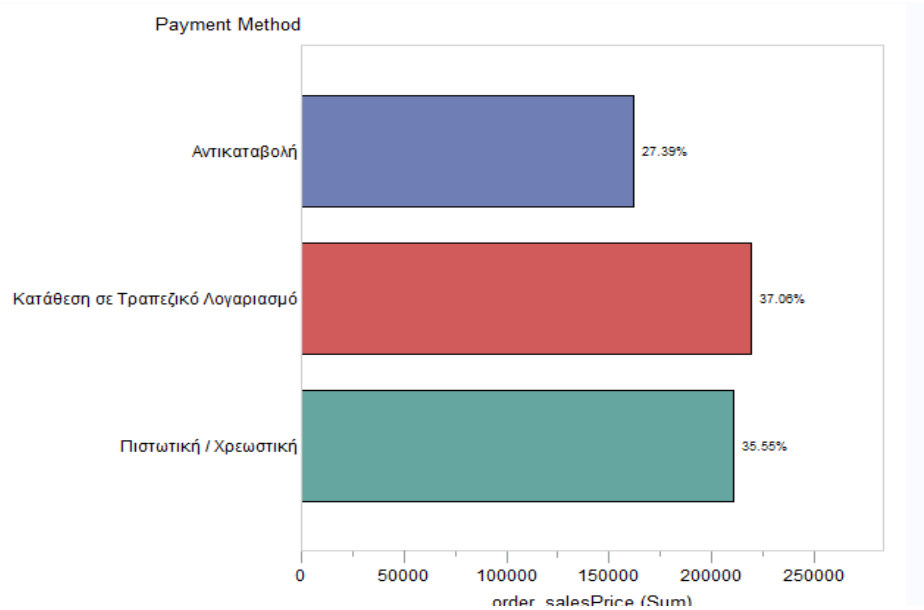


Chart 8 - Πωλήσεις με βάση τον τρόπο πληρωμής

Κατόπιν περαιτέρω διερεύνησης και αναλύοντας τον τρόπο πληρωμών των παραγγελιών σε βάθος χρόνου, διαπιστώνεται ότι οι πελάτες από το 2015 έως το 2019 επιλέγουν την αντικαταβολή συντριπτικά σε σχέση με τις άλλες δυνατότητες. Από το 2020 και μετά η συμπεριφορά των καταναλωτών μεταβάλλεται τελείως, προχωρώντας σε χρήση των τραπεζικών πιστωτικών καρτών ή πληρωμή με κατάθεση σε τραπεζικό λογαριασμό.

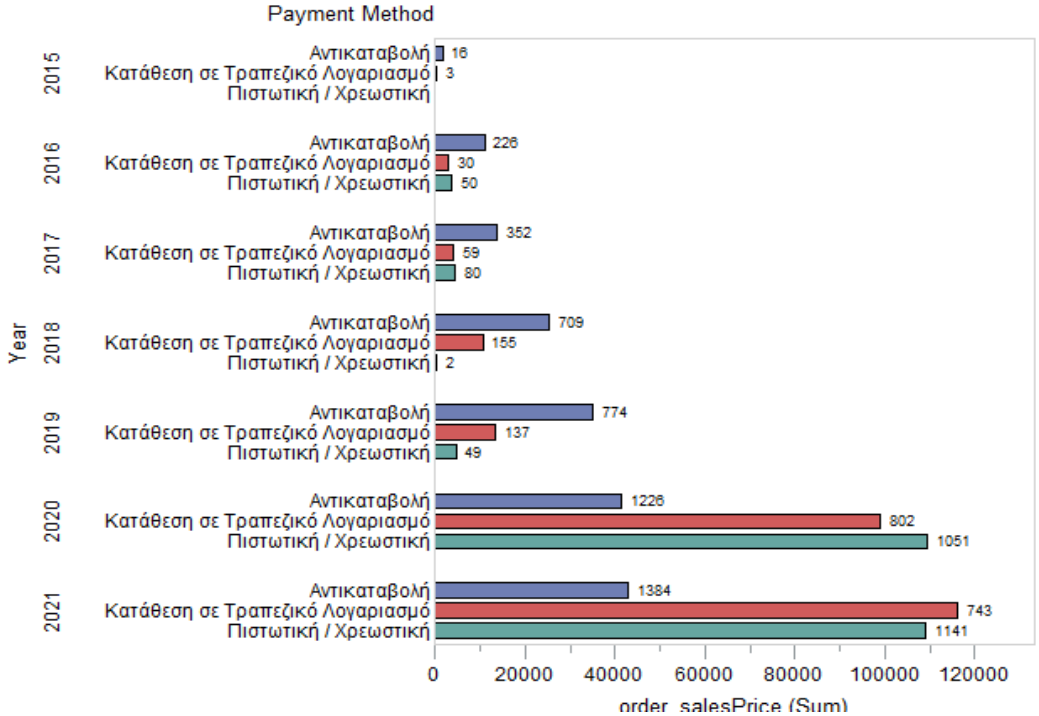


Chart 9 - Πωλήσεις ανά τρόπο πληρωμής ανά έτος

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

6

Εφαρμογή RFM Analysis - Market Basket Analysis και Αξιολόγηση αποτελεσμάτων

6.1 RFM Model

6.1.1 Παράμετροι αξιολόγησης

Όπως περιεγράφηκε στη θεωρία στο Κεφάλαιο 3, για να είναι εφικτή η εκτέλεση του RFM Model Analysis, πρέπει να προκύψουν τα τρία χαρακτηριστικά των πελατών: το Recency, το Frequency και το Monetary value.

Για την παραγωγή αυτών των στοιχείων θα χρησιμοποιηθούν τα datasets. Συγκεκριμένα, για την παραγωγή του recency, η πληροφορία θα αντληθεί από το αρχείο Orders payments analysis υπολογίζοντας του μήνες από την τελευταία αγορά που πραγματοποίησε ο πελάτης. Στη συνέχεια, για το Frequency θα καταμετρηθούν οι μοναδικές παραγγελίες που έχει πραγματοποιήσει ο κάθε πελάτης. Τέλος, το Monetary θα υπολογιστεί προσθέτοντας το κόστος πώλησης των παραγγελιών ανά πελάτη.

Τα Metrics αυτά θα παραχθούν από την εφαρμογή SAS Enterprise Guide και το παραγόμενο Dataset θα αναλυθεί μέσω της εφαρμογής SAS Enterprise Miner. Αρχικά, θα πρέπει το παραγόμενο Dataset να αναλυθεί και να εξεταστούν ακραίες τιμές που πιθανόν να πρέπει να εξαιρεθούν.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

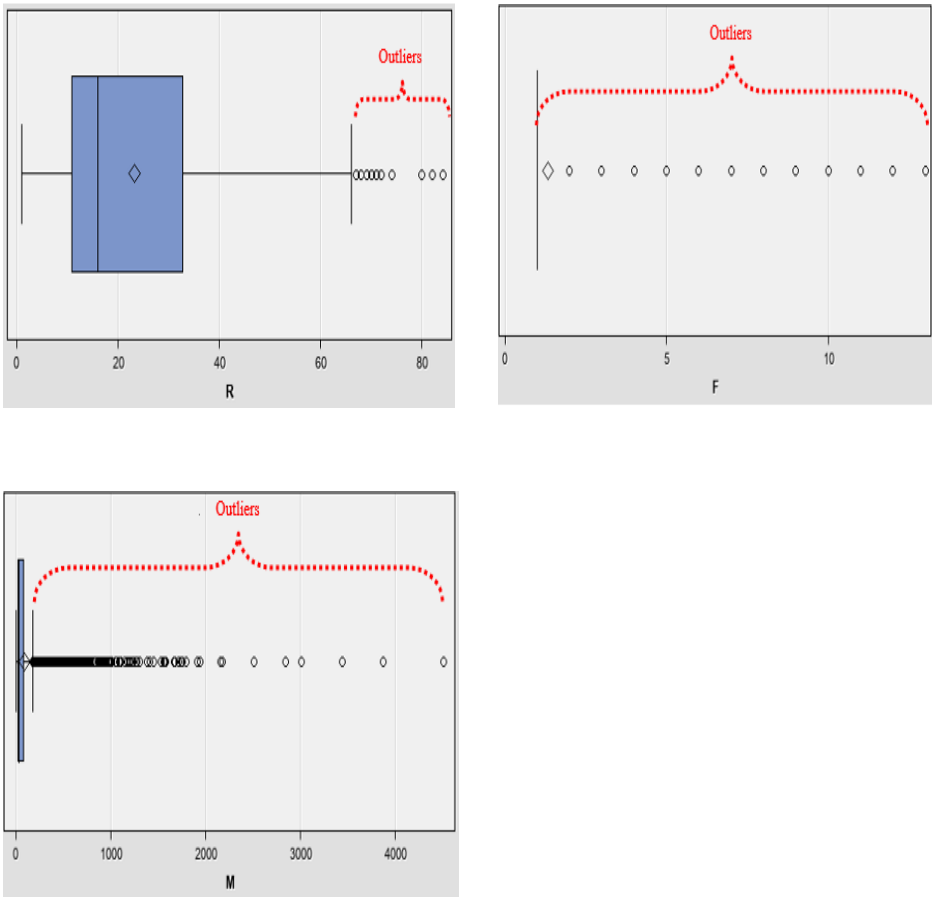


Chart 10 - Box Plots for Recency, Frequency and Monetary

Αναλύοντας τα box plots ανά μεταβλητή εντοπίζονται τα Outliers τα οποία πρέπει να αφαιρεθούν. Για το R οι τιμές που θα αφαιρεθούν είναι από 67 και πάνω. Για το F διαπιστώνεται ότι η πλειονότητα των πελατών έχει αγοράσει μόνο μία φορά από το κατάστημα και για αυτό είναι τόσο μεγάλο το διάστημα. Για το Monetary οι τιμές που πρέπει να αφαιρεθούν είναι από 180 και πάνω.

Μετά την αφαίρεση των ακραίων τιμών η απεικόνιση του box plot για το recency διορθώθηκε, αλλά δεν συμβαίνει το ίδιο με το Monetary. Κατά συνέπεια, θα πρέπει να προχωρήσει η έρευνα με Log transformation.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

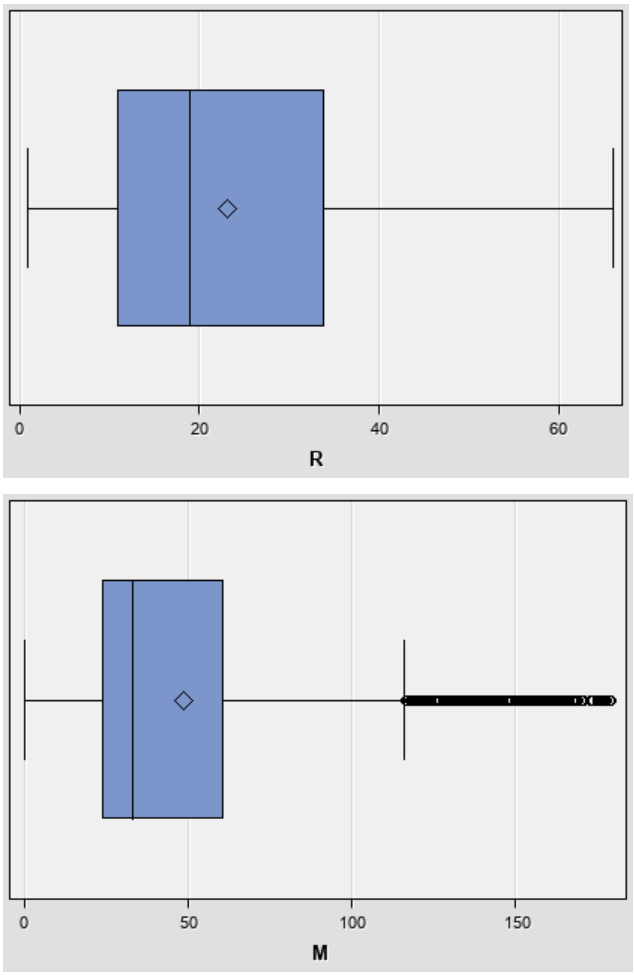


Chart 11 - Box Plots for Recency και Monetery μετά την αφαίρεση ακραίων τιμών

Μετά το variable transformation το μοντέλο παρέδωσε 5 Clusters.

Segment Id	Frequency of Cluster	R	F	M
1	494	19.93	1.11	67.54
2	3062	15.28	0.69	33.25
3	1403	45.50	0.70	32.44
4	688	16.20	0.70	121.43
5	135	13.10	1.58	128.26

Πίνακας 10 - Segments of RFM

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

6.1.2 Επεξήγηση των Clusters

	MIN	MEAN	MAX
Segment Id 1			
Recency	1.00	19.93	60.00
Frequency	1.10	1.11	1.61
Monetary	3.80	67.54	156.76
Segment Id 2			
Recency	1.00	15.28	30.00
Frequency	0.69	0.69	0.69
Monetary	0.10	33.25	77.20
Segment Id 3			
Recency	31.00	45.50	66.00
Frequency	0.69	0.70	1.10
Monetary	0.81	32.44	122.53
Segment Id 4			
Recency	1.00	16.20	62.00
Frequency	0.69	0.70	1.10
Monetary	77.00	121.43	179.34
Segment Id 5			
Recency	1.00	13.10	60.00
Frequency	1.39	1.58	2.30
Monetary	67.80	128.26	179.18

Πίνακας 11 - Τιμές RFM ανά Segment

Cluster 1

Οι πελάτες που περιέχονται στο Cluster ξοδεύουν κατά μέσο όρο 67€ σε αγορές. Θεωρητικά εμφανίζονται αρκετά συχνά για παραγγελίες, αλλά η τελευταία φορά που αγόρασαν προϊόντα ήταν πριν 2 έτη.

Cluster 2

Οι συγκεκριμένοι πελάτες αν και επισκέφθηκαν πρόσφατα το κατάστημα, ξοδεύουν περίπου 33€, αλλά δεν εμφανίζονται τόσο συχνά στο κατάστημα.

Cluster 3

Οι πελάτες που περιέχονται στο Cluster 3 ξοδεύουν κατά μέσο όρο 32€ σε αγορές, όμως η τελευταία φορά που αγόρασαν ήταν πριν πολύ καιρό και η συχνότητα επισκέψεών τους είναι μικρή.

Cluster 4

Οι πελάτες στο συγκεκριμένο Cluster ξοδεύουν αρκετά χρήματα στις αγορές τους με μέσο όρο τα 121€. Δεν εμφανίζονται συχνά, αλλά πραγματοποίησαν πρόσφατα αγορές.

Cluster 5

Πρόκειται για την καλύτερη ομάδα πελατών, καθώς ξοδεύουν τα περισσότερα χρήματα με μέσο όρο 128€, πολύ πρόσφατα αγόρασαν προϊόντα και η συχνότητά τους είναι μεγάλη.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

6.1.3 Segment Profiling

Με βάση τα παραπάνω και προκειμένου να χαρακτηριστούν τα Segments, ακολουθείται η μέθοδος σύγκρισής των μέσων με τους μέσους του συνόλου. Έτσι, από τον παρακάτω πίνακα συγκρίνονται όλες οι τιμές με βάση την overall τιμή για κάθε Cluster και καταλήγουμε στις ονομασίες τους: Promising, New Customers, Lost, Need Attention και High Value.

Segment Id	R	F	M	Segement Description
1	19.93	1.11	67.54	Promising
2	15.28	0.69	33.25	New Customers
3	45.50	0.70	32.44	Lost
4	16.20	0.70	121.43	Need Attention
5	13.10	1.58	128.26	High Value
Overall	**23.07**	**1.16**	**48.69**	

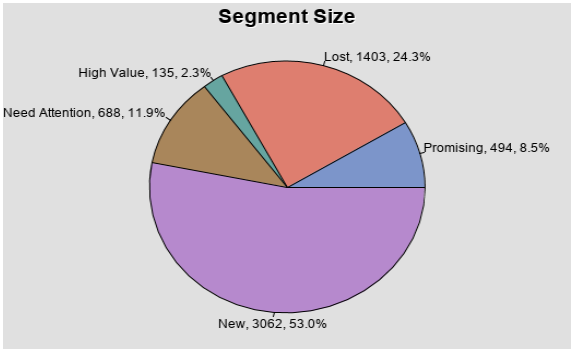
Πίνακας 12 - RFM Segements Description

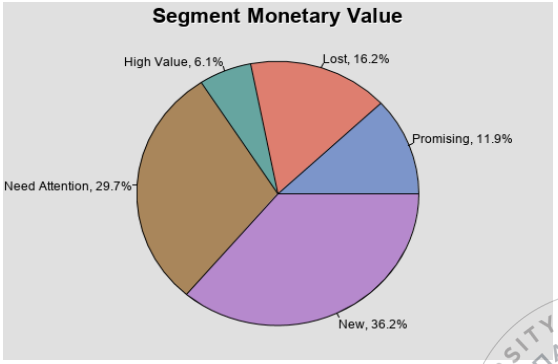
Ο παραπάνω πίνακας απεικονίζει μια συνοπτική παρουσίαση του πελατολογίου, τόσο για τους άμεσα ενδιαφερόμενους της επιχείρησης, όσο και για τον ίδιο τον ερευνητή.

6.1.1 Αποτελέσματα

Αφού έχει τρέξει το μοντέλο και ολοκληρωθηκε η παραγωγή των Clusters και έπειτα η ονοματοδοσία τους, προκειμένου να γίνουν αντιληπτά τα μεγέθη, παραθέτωνται δύο Pie charts με την ποσοτικοποίηση των Clusters στο πελατολόγιο και την απεικόνιση των Clusters με βάσει το Monetary, ώστε να αντιστοιχηθούν τα Segmnets στον τζίρο της εταιρίας.

Έτσι, προκύπτουν τα παρακάτω pie Charts και τα ακόλουθα συμπεράσματα:





ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

1. Οι **Promising** (υποσχόμενοι πελάτες), αν και αποτελούν μικρό μέρος του πελατολογίου με ποσοστό 8,5%, συνεισφέρουν στο 12% του τζίρου. Οι συγκεκριμένοι πελάτες πρέπει να αποκτήσουν κίνητρο για να παραγγέλνουν πιο σύχνα και να μην αποσουσιάζουν μεγάλα διαστήματα από την τελευταία φορά που αγόρασαν. Μια πιθανή καμπάνια με αποστολή εκπτωτικών κουπονιών μέσω ηλεκτρονικής αλληλογραφίας θα βελτιώσει τα αποτελέσματα.

2. Οι **New customers** αποτελούν ένα σημαντικό μέρος, αφού αποτελούν το 36% του τζίρου, παρόλο που ο μέσος όρος είναι μικρός. Κατέχουν το μεγαλύτερο μέρος του συνόλου των πελατών με 53%. Με τις κατάλληλες προωθητικές ενέργειες θα αυξηθεί η συχνότητα παραγγελιών και κατά συνέπεια, θα αυξηθεί ο μέσος όρος του κόστους αγορών. Ένα πρόγραμμα επιβράβευσης θα αποτελέσει κίνητρο για την συνέχιση των αγορών και περαιτέρω αύξηση των εσόδων.

3. Οι **Lost** πελάτες κατέχουν ένα σημαντικό ποσοστό στο πελατολόγιο, κάτι το οποίο δεν πρέπει να αγνοηθεί. Ίσως το πιο σημαντικό Pool για το οποίο πρέπει να οργανωθούν καμπάνιες προωθητικών ενεργειών για να επιστρέψουν οι πελάτες στην επιχείρηση.

4. Οι **Need Attention** είναι πελάτες που συμμετέχουν σχεδόν στο 1/3 του τζίρου του οργανισμού. Παρόλο που είναι μικρό ποσοστό στο σύνολο των πελατών, αποτελεί πολύ σημαντικό κομμάτι των εσόδων. Πρέπει να συσταθούν κατάλληλες προωθητικές ενέργειες με προσφορές και προτάσεις σε προιόντα που τους ενδιαφέρουν για να αυξηθεί η συχνότητα παραγγελιών.

5. Οι **High Value** πελάτες είναι το μικρότερο Cluster με 2,3% επί του συνόλου και 6,1% επί του τζίρου της επιχείρησης. Μέσω προγράμματος επιβράβευσης και με σκοπό να κατηγοριοποιηθούν ως VIP μέλη, θα ενισχύσουν την πελατειακή σχέση με την επιχείρηση, ώστε να διατηρηθούν στο καλό πελατολόγιο της.

6.1.2 Σύνοψη συμπερασμάτων αξιολόγησης

Από τα αποτελέσματα της ανάλυσης δημιουργείται έδαφος για ενέργειες προς το πελατολόγιο, οι οποίες θα βελτιώσουν σε μεγάλο βαθμό τα έσοδα της εταιρίας. Η πρόταση για την υιοθέτηση ενός Loyalty program θα προσθέσει τα παρακάτω οφέλη:

- Θα ενισχύσει την καλή επικοινωνία της εταιρίας με τους πελάτες της.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

- Θα δημιουργήσει την αίσθηση ότι οι πελάτες αποτελούν μέρος της επιχείρησης και συνακόλουθα, μεγαλύτερη οικειότητα.
- Θα προσελκύσει περισσότερους πελάτες στο πελατειακό group 18-30, το οποίο είναι περισσότερο εξοικειωμένοι με την τεχνολογία.
- Επιπλέον, μέσω του προγράμματος θα ενισχυθεί η καλή σχέση με τους High Value πελάτες, καθώς θα ενημερωθούν πως η επιχείρηση τους θεωρεί VIP members.
- Θα δημιουργηθεί δίαυλος επικοινωνίας για άμεσες προωθητικές ενέργειες.
- Θα επιχειρηθεί η επανεκκίνηση των πελατών Lost που πλέον αγοράζουν από άλλους οργανισμούς.

6.2 Market Basket Analysis

Η μέθοδος αποτελεί μια από τις πιο γνωστές στον χώρο των Data Analytics. Η εφαρμογή της επιτρέπει στους εμπορικούς οργανισμούς να λάβουν εκείνες τις αποφάσεις που θα συμβάλλουν στην αύξηση των πωλήσεων και των κερδών.

Πρώτα από όλα για την ανάλυση απαιτείται ένα dataset συναλλαγών, όπου ανά απόδειξη απεικονίζει διακριτά τα προϊόντα (SKU). Για τις ανάγκες της διπλωματικής χρησιμοποιήθηκε το αρχείο με ονομασία “Order Items”. Αρχικά, θα αξιολογηθεί το αρχείο, ώστε να αποφασιστεί εάν πρέπει να μειωθεί το δείγμα για την έρευνα.

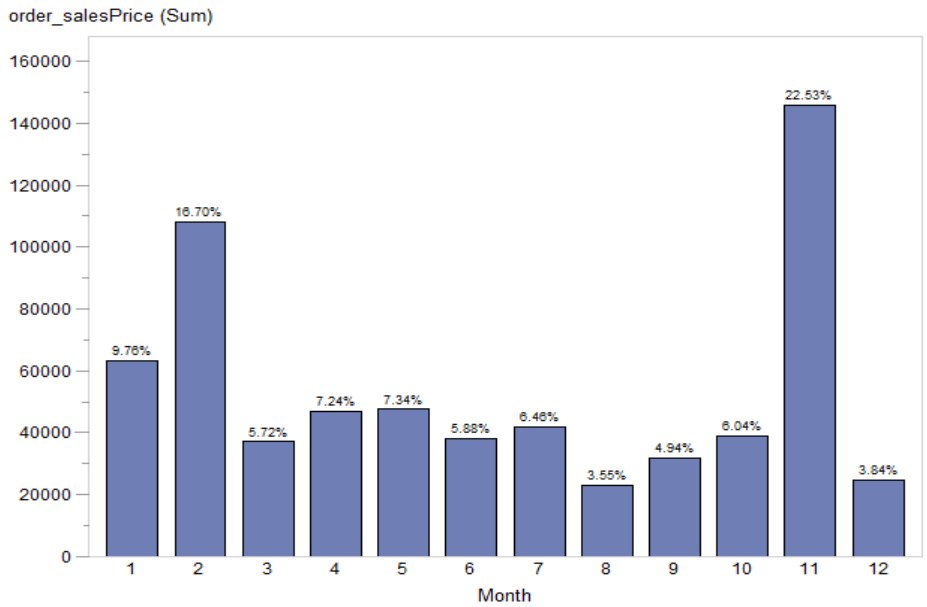


Chart 12 - % πωλήσεων ανά μήνα μέσα στο έτος

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Διαπιστώνεται ότι οι πωλήσεις αυξάνονται τους μήνες Νοέμβριο και Φεβρουάριο, κάτι το οποίο είναι απόλυτα λογικό, καθώς αφορά μια επιχείρηση που εμπορεύεται εποχιακά είδη. Συνεπώς, τους μήνες πριν τα Χριστούγεννα και κατά την περίοδο των αποκριών – συνήθως εντός Φεβρουαρίου – οι πωλήσεις είναι πολύ μεγαλύτερες. Αίσθηση προκαλεί πως ο Δεκέμβριος είναι ο 2ος μήνας με τις λιγότερες παραγγελίες. Επειδή το data set περιλαμβάνει παραγγελίες από το 2015, θα περιορίσουμε το χρονικό διάστημα από την 01/01/2021 και μετά.

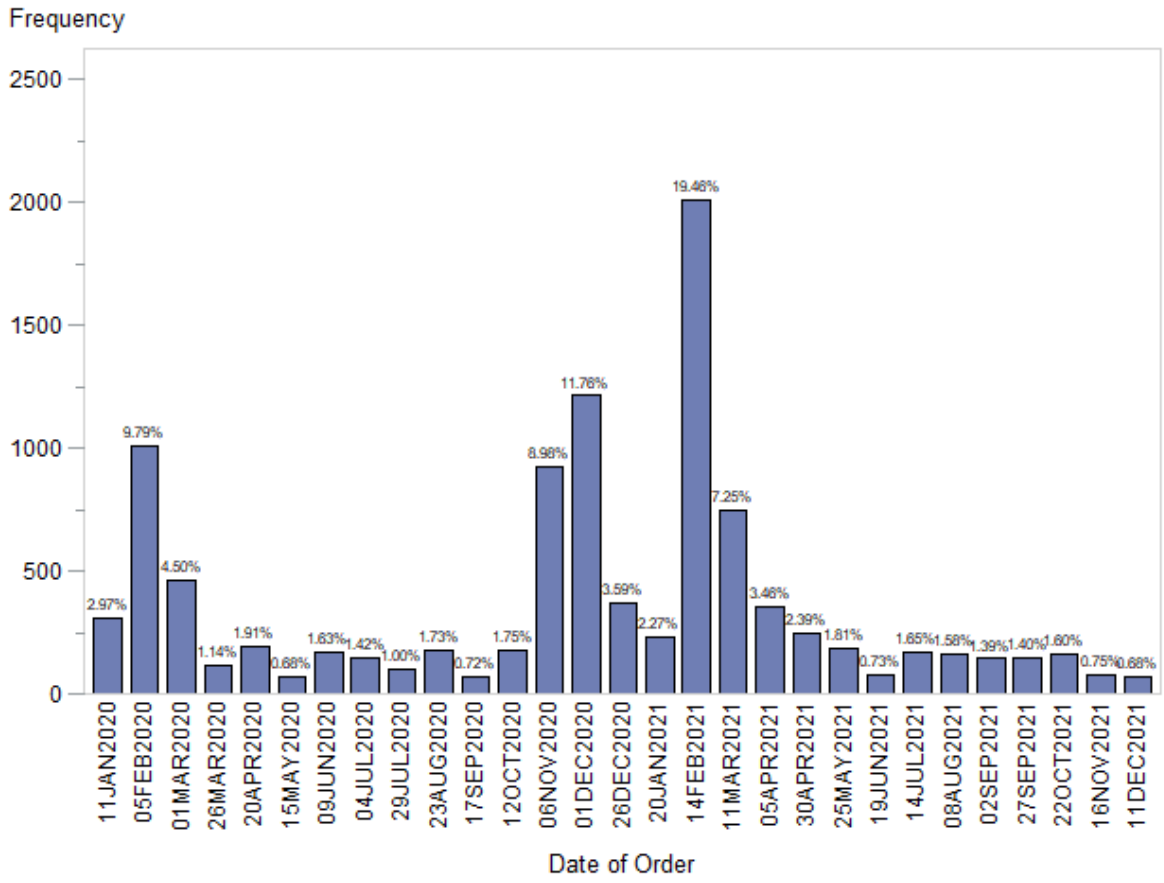


Chart 13 - % πωλήσεων ανά μήνα 2020-2021

Αν και με αυτόν τον τρόπο μειώθηκε το χρονικό διάστημα αξιολόγησης, η γενική εικόνα παραμένει, καθώς οι μέγιστες παραγγελίες ολοκληρώνονται στους μήνες Φεβρουάριο και Νοεμβρίο. Τα αποτελέσματα που προκύπτουν επηρεάζονται από τις συνθήκες λόγω της πανδημίας, όπου για μεγάλα χρονικά διαστήματα οι πολίτες αναγκάστηκαν να παραμείνουν σε κατ΄ οίκον περιορισμό και οι περισσότερες επιχειρήσεις δε λειτουργούσαν. Κατά συνέπεια, τον Νοέμβριο του 2020 οι παραγγελίες αυξήθηκαν, καθώς τότε είχαν ξεκινήσει μέτρα περιορισμού των μετακινήσεων και το καταναλωτικό κοινό αφιέρωνε πολύ περισσότερο χρόνο στην πλοήγηση στο διαδίκτυο για τις

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

αγορές που ήθελε να πραγματοποιήσει. Αξιοσημείωτο είναι το γεγονός ότι ο αντίστοιχος Νοέμβριος του 2021 παρουσιάζει μεγάλη πτώση πωλήσεων σε σχέση με το 2020.

Όσον αφορά στα προϊόντα που αγοράζουν οι πελάτες, στον παρακάτω πίνακα απεικονίζονται τα 30 πιο συχνά αγορασμένα προϊόντα. Κατά κύριο λόγο επιλέγονται Χριστουγεννιάτικα είδη και αποκριάτικα είδη.

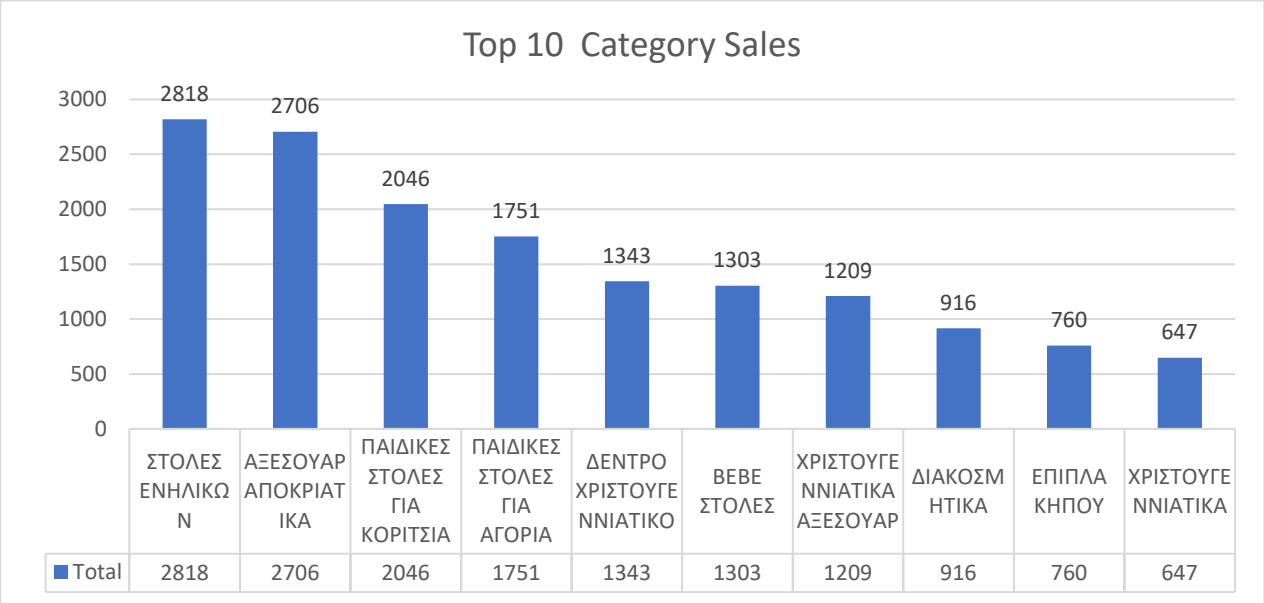


Chart 14 - Πωλήσεις ανά κατηγορία προϊόντος

6.2.1 Μεθοδολογία MBA

Για να μπορέσουμε να διεξάγουμε την ανάλυση θα πρέπει να ακολουθήσουμε τα βήματα της μεθοδολογίας για την εκτέλεσή της. Θα ακολουθήσει η περιγραφή πέντε βημάτων που καλύπτουν όλη την ερευνητική διαδικασία. Αρχικά, (βήμα 1) πρέπει να αξιολογηθεί εάν το MBA είναι κατάλληλο για τις συναλλαγές του Οργανισμού που αναλύεται. Στη συνέχεια, (βήμα 2) πρέπει να αποφασιστεί ποιες συναλλαγές ικανοποιούν τις προδιαγραφές της έρευνας. Επόμενη ενέργεια (βήμα 3) είναι η συλλογή των δεδομένων και ο έλεγχος για το εάν πληρούνται κάποια κριτήρια για το μοντέλο. Πλέον (βήμα 4) είναι εφικτό να τρέξει το μοντέλο και τέλος, (βήμα 5) να ερμηνευτούν τα αποτελέσματα (Aguinis, Forcum, & Joo, 2013).

Βήμα 1: Αξιολόγηση εάν το MBA είναι κατάλληλο για την έρευνα

Το MBA χρησιμοποιείται για εύρεση κανόνων συσχετίσεων, για το λόγο αυτό βασίζεται σε δεδομένα με δυαδική μορφή και με κατηγορικές μεταβλητές, οι οποίες πρέπει να έχουν σημασία και

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ενδιαφέρον μεταξύ τους. Οι βάσεις δεδομένων συναλλαγών εξυπηρετούν απόλυτα το σκοπό αυτό και ο στόχος μέσα από την ανάλυση να παραχθούν κανόνες συσχέτισης, επιβεβαιώνει ότι το μοντέλο είναι κατάλληλο. Ως εκ τούτου, για την εφαρμογή του μοντέλου στον εμπορικό οργανισμό που αξιολογείται, το MBA κρίνεται πως είναι κατάλληλο.

Βήμα 2: Προσδιορισμός των συναλλαγών που θα χρησιμοποιηθούν

Όταν ένας πελάτης αγοράσει προϊόντα ή υπηρεσίες από φυσικό ή ηλεκτρονικό κατάστημα, λαμβάνει μια απόδειξη με έναν μοναδικό κωδικό συναλλαγής, η οποία περιέχει όλα τα είδη που πλήρωσε. Σε μια βάση δεδομένων συναλλαγών δημιουργείται μια εγγραφή για κάθε προϊόν αγοράς, όπου αναφέρονται ο αριθμός της απόδειξης, ο κωδικός προϊόντος, η περιγραφή προϊόντος, πιθανόν το κόστος, η ποσότητα και η ημερομηνία αγοράς. Το Dataset “Order Items” του υπό ανάλυση οργανισμού είναι το κατάλληλο για το MBA.

Βήμα 3: Συλλογή δεδομένων και έλεγχος απαιτήσεων MBA

Ο ερευνητής προκειμένου να εκτελέσει τις αναλύσεις του πρέπει να επεξεργαστεί τα κατάλληλα δεδομένα. Χρειάζεται ένας μεγάλος όγκος συναλλαγών για να καταλήξει σε ορθά συμπεράσματα. Το data set που θα χρησιμοποιηθεί στην ανάλυση έχει περιοριστεί στη χρονική διάρκεια από 01/01/2020 έως το τέλος του 2021 με 5.870 συναλλαγές.

Βήμα 4: Εκτέλεση ανάλυσης

Αφού τα data που έχει στη διάθεσή του ο ερευνητής ικανοποιούν όλες τις συνθήκες για να προχωρήσει η ανάλυση, θα χρησιμοποιηθεί το Enterprise Miner της εταιρίας SAS για την παραγωγή των κανόνων συσχέτισης. Παρακάτω θα αξιολογηθούν τα αποτελέσματα που θα παραχθούν και στο κεφάλαιο 7 αποτυπώνονται περισσότερες τεχνικές λεπτομέρειες της εφαρμογής.

Βήμα 5: Ερμηνεία αποτελεσμάτων

Το τελευταίο βήμα είναι η ερμηνεία των αποτελεσμάτων. Από την εκτέλεση του μοντέλου θα παραχθούν κανόνες συσχέτισης. Η ανάλυση τους θα επιτρέψει στον ερευνητή να καταλήξει σε συμπεράσματα και προτάσεις προς τον οργανισμό. Οι παράμετροι αξιολόγησης που θα χρησιμοποιηθούν είναι το Lift, το Support και το Confidence.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

6.2.2 Αποτελέσματα

Αρχικά, η πρώτη προσπάθεια πραγματοποιήθηκε με Target την μεταβλητή Item Description, μέγιστο αριθμό items 4 στον κανόνα, το minimum confidence στο 10% και το Support στο 5%. Σύμφωνα με έρευνες της SAS οι προηγούμενες παραμετροποιήσεις θεωρούνται ως ικανοποιητικές, αν και δεν θεωρείται κάποιο set ως ιδανικό.

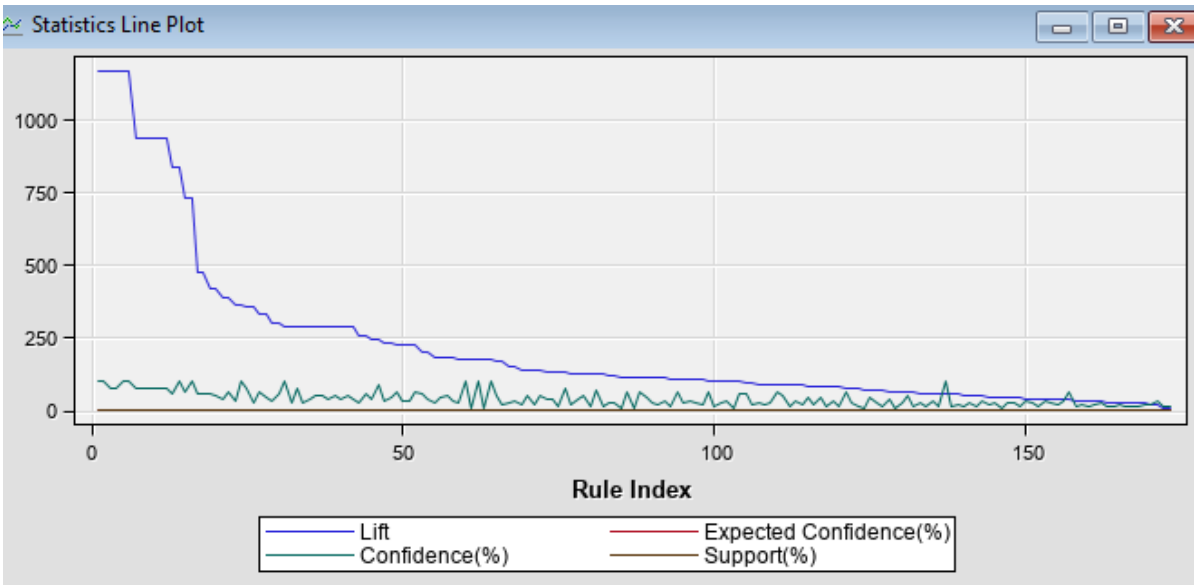


Chart 15 - Διάγραμμα παραγόμενων κανόνων

Από το διάγραμμα στην παραπάνω εικόνα διαπιστώνεται ότι έχουν δημιουργηθεί περισσότεροι από 150 κανόνες. Για την ακρίβεια οι κανόνες που δημιούργησε το λογισμικό είναι 173.

Αναλυτικότερα στον πίνακα των κανόνων απεικονίζονται οι 25 πρώτοι κανόνες συσχέτισης ταξινομημένοι με φθίνουσα σειρά βάσει του Lift. Προκύπτει ότι οι περισσότεροι κανόνες αφορούν σε προϊόντα για πάρτι γενεθλίων.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

RULE A/A	EXP_CONF	CONF	SUPPORT	LIFT	COUNT	RULE
1	0.15	38.10	0.14	248.47	8	ΘΕΟΣ ΑΡΗΣ ==> ΑΣΠΙΔΑ ΜΟΝΟΜΑΧΟΥ - 57ΕΚ.
2	0.36	88.89	0.14	248.47	8	ΑΣΠΙΔΑ ΜΟΝΟΜΑΧΟΥ - 57ΕΚ. ==> ΘΕΟΣ ΑΡΗΣ
3	0.29	58.33	0.12	201.42	7	ΜΠΑΛΑ ΡΟΖ ΒΕΛΟΥΔΟ ==> ΜΠΑΛΑ ΜΠΟΡΝΤΩ ΒΕΛΟΥΔΟ
4	0.20	41.18	0.12	201.42	7	ΜΠΑΛΑ ΜΠΟΡΝΤΩ ΒΕΛΟΥΔΟ ==> ΜΠΑΛΑ ΡΟΖ ΒΕΛΟΥΔΟ
5	0.17	30.30	0.17	177.88	10	ΣΠΑΘΙ ΠΕΙΡΑΤΗ ==> ΜΑΤΙ ΠΕΙΡΑΤΗ ΥΦΑΣΜΑΤΙΝΟ
6	0.56	100.00	0.17	177.88	10	ΜΑΤΙ ΠΕΙΡΑΤΗ ΥΦΑΣΜΑΤΙΝΟ ==> ΣΠΑΘΙ ΠΕΙΡΑΤΗ
7	0.29	38.89	0.12	134.28	7	Χριστουγενιάτικη Μπάλα με σχέδια - ΣΕΤ 6ΤΜΧ ==> Χριστουγενιάτικη Μπάλα Glitter
8	0.31	41.18	0.12	134.28	7	Χριστουγενιάτικη Μπάλα Glitter - ΣΕΤ 6ΤΜΧ ==> Χριστουγενιάτικη Μπάλα με σχέδια
9	0.14	18.18	0.10	133.41	6	ΣΠΑΘΙ ΠΕΙΡΑΤΗ ==> ΚΟΥΡΣΑΡΟΣ ΤΗΣ ΚΑΡΑΪΒΙΚΗΣ
10	0.56	75.00	0.10	133.41	6	ΚΟΥΡΣΑΡΟΣ ΤΗΣ ΚΑΡΑΪΒΙΚΗΣ ==> ΣΠΑΘΙ ΠΕΙΡΑΤΗ

Πίνακας 13 - 10 κανόνες με φθίνουσα σειρά βάσει Lift

Αναλύοντας τον κανόνα 1 διαπιστώνεται ότι όταν ο πελάτης έχει στο καλάθι του το προϊόν «Μπαλόνι αριθ 7 Rose Gold», έχει πιθανότητα 1174 να αγοράσει και το προϊόν «Γιρλάντα Χαρούμενα Γενέθλια». Το Support είναι ιδιαίτερα χαμηλό, αποτέλεσμα που οφείλεται στο γεγονός ότι τα δυο παραπάνω προϊόντα εντοπίζονται σε 5 συναλλαγές, ενώ το σύνολο των συναλλαγών είναι πολύ μεγάλο.

Διαπιστώνεται ότι λόγω του μεγάλου αριθμού κωδικών προϊόντων, οι συσχετίσεις που παράγονται δε βοηθούν στο να προκύψουν συμπεράσματα.

Για τον παραπάνω λόγο, θα εκτελεστεί εκ νέου το πείραμα με Target μεταβλητή πλέον την κατηγορία των προϊόντων. Οι παράμετροι για το minimum confidence στο 10% και το Support στο 5% παραμένουν.

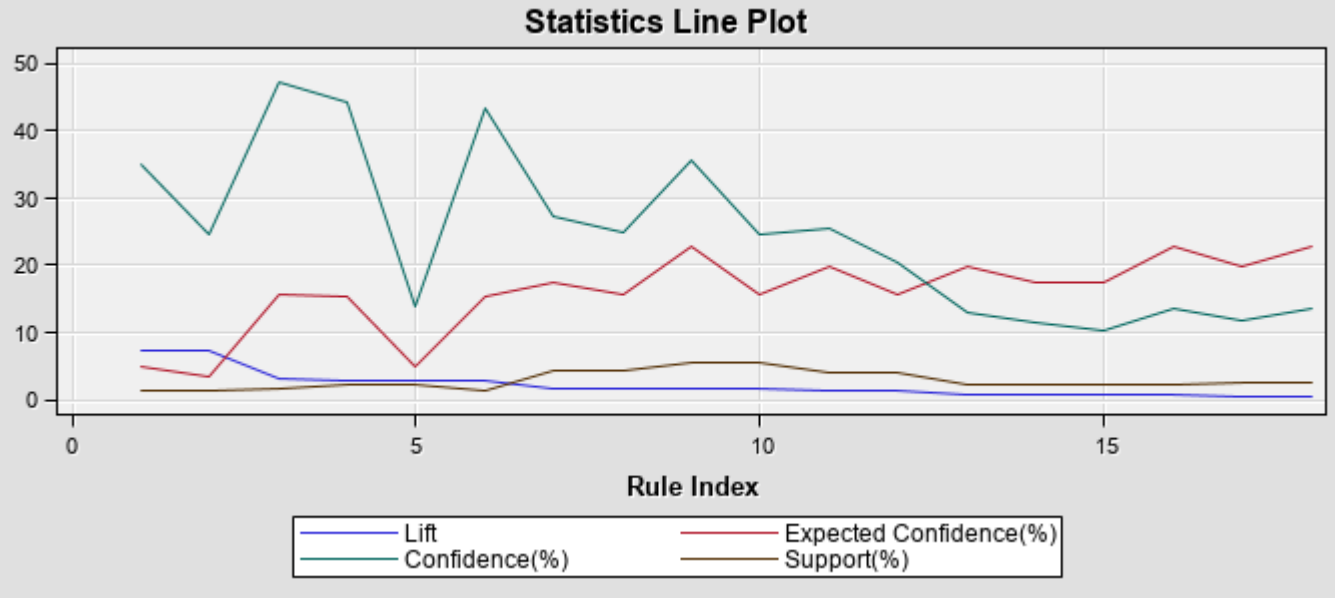


Chart 16 - Διάγραμμα κανόνων μετά αύξηση Support & Confidence

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Το λογισμικά παρήγαγε τώρα 18 κανόνες. Στον παρακάτω πίνακα συγκεντρώνονται οι top 10 κανόνες με το υψηλότερο Lift.

RUL A/A	EXP CONF	CONF	SUPPORT	LIFT	COUNT	RULE
1	4.84	34.97	1.19	7.23	100	ΧΡΙΣΤΟΥΓΕΝΝΙΑΤΙΚΑ ΑΞΕΣΟΥΑΡ ==> ΦΩΤΑΚΙΑ-LED
2	3.41	24.63	1.19	7.23	100	ΦΩΤΑΚΙΑ-LED ==> ΧΡΙΣΤΟΥΓΕΝΝΙΑΤΙΚΑ ΑΞΕΣΟΥΑΡ
3	15.71	47.19	1.50	3.00	126	ΑΠΟΚΡΙΑΤΙΚΕΣ ΠΕΡΟΥΚΕΣ ==> ΑΞΕΣΟΥΑΡ ΑΠΟΚΡΙΑΤΙΚΑ
4	15.33	44.09	2.13	2.88	179	ΦΩΤΑΚΙΑ-LED ==> ΔΕΝΤΡΟ ΧΡΙΣΤΟΥΓΕΝΝΙΑΤΙΚΟ
5	4.84	13.91	2.13	2.88	179	ΔΕΝΤΡΟ ΧΡΙΣΤΟΥΓΕΝΝΙΑΤΙΚΟ ==> ΦΩΤΑΚΙΑ-LED
6	15.33	43.36	1.48	2.83	124	ΧΡΙΣΤΟΥΓΕΝΝΙΑΤΙΚΑ ΑΞΕΣΟΥΑΡ ==> ΔΕΝΤΡΟ ΧΡΙΣΤΟΥΓΕΝΝΙΑΤΙΚΟ
7	17.28	27.29	4.29	1.58	360	ΑΞΕΣΟΥΑΡ ΑΠΟΚΡΙΑΤΙΚΑ ==> ΠΑΙΔΙΚΕΣ ΣΤΟΛΕΣ ΓΙΑ ΑΓΟΡΙΑ
8	15.71	24.81	4.29	1.58	360	ΠΑΙΔΙΚΕΣ ΣΤΟΛΕΣ ΓΙΑ ΑΓΟΡΙΑ ==> ΑΞΕΣΟΥΑΡ ΑΠΟΚΡΙΑΤΙΚΑ
9	22.76	35.48	5.57	1.56	468	ΑΞΕΣΟΥΑΡ ΑΠΟΚΡΙΑΤΙΚΑ ==> ΣΤΟΛΕΣ ΕΝΗΛΙΚΩΝ
10	15.71	24.49	5.57	1.56	468	ΣΤΟΛΕΣ ΕΝΗΛΙΚΩΝ ==> ΑΞΕΣΟΥΑΡ ΑΠΟΚΡΙΑΤΙΚΑ

Πίνακας 14 - Κανόνες μετά την αύξηση Support & Confidence

Επεξήγηση κανόνων

Κανόνας 1: Για τον πρώτο κανόνα το Lift είναι 7.23 για την περίπτωση που ο πελάτης θα αγοράσει «ΦΩΤΑΚΙΑ-LED», εφόσον έχει ήδη τοποθετήσει στο καλάθι του την αποκριάτικη στολή «ΧΡΙΣΤΟΥΓΕΝΝΙΑΤΙΚΑ ΑΞΕΣΟΥΑΡ». Το Support είναι 1,19% και το Confidence 34,97%. Άρα είναι 7.23 φορές πιθανότερο ένας πελάτης που αγοράζει προϊόντα με Χριστουγεννιάτικα αξεσουάρ, να αγοράσει και Φωτάκια LED.

Κανόνας 2: Περιγράφεται ο κανόνας 1 με την αντίθετη φορά, καθώς στο αριστερό μέρος βρίσκονται τα «ΦΩΤΑΚΙΑ-LED» και στο δεξί τα «ΧΡΙΣΤΟΥΓΕΝΝΙΑΤΙΚΑ ΑΞΕΣΟΥΑΡ». Συνεπώς, αν ο πελάτης επιλέξει τα φωτάκια, τότε θα επιλέξει και άλλα Χριστουγεννιάτικα είδη με πιθανότητα 7.23.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Κανόνας 3: Για τον τρίτο κανόνα το Lift είναι 3 και όταν ο πελάτης θα επιλέξει «ΑΠΟΚΡΙΑΤΙΚΕΣ ΠΕΡΟΥΚΕΣ», θα επιλέξει και «ΑΞΕΣΟΥΑΡ ΑΠΟΚΡΙΑΤΙΚΑ». Η αύξηση του Confidence 47.19% στον κανόνα σημαίνει ότι ο συνδυασμός των δυο κατηγοριών στα καλάθια αγορών είναι σχεδόν 50% στις συναλλαγές που περιέχουν αποκριάτικες περούκες και μεταφράζεται ότι ο πελάτης με τις κατάλληλες προτάσεις θα προσθέσει επιπλέον αντικείμενα στις επιλογές του.

6.2.3 Σύνοψη συμπερασμάτων αξιολόγησης

Από τα αποτελέσματα του Market Basket Analysis προκύπτει ότι οι πελάτες συνδυάζουν προϊόντα ανά συσχετιζόμενη κατηγορία. Επικρατεί η κατηγορία των αποκριάτικων ειδών στο σύνολο των συναλλαγών και ακολουθεί η κατηγορία των Χριστουγεννιάτικων ειδών, παρόλο που τα Χριστουγεννιάτικα είχαν υψηλότερο Lift. Η επιχείρηση μπορεί στο site να υλοποιήσει λειτουργικότητες προτάσεων προϊόντων προς τον πελάτη, που σχετίζονται με τα προϊόντα που έχει ήδη στο καλάθι του. Επίσης, κυρίως για την εποχή των Αποκριών το τμήμα πωλήσεων μπορεί να υιοθετήσει συνδυαστικές προσφορές σε προϊόντα που θα επιλέξει ο πελάτης να τοποθετήσει στο καλάθι του.

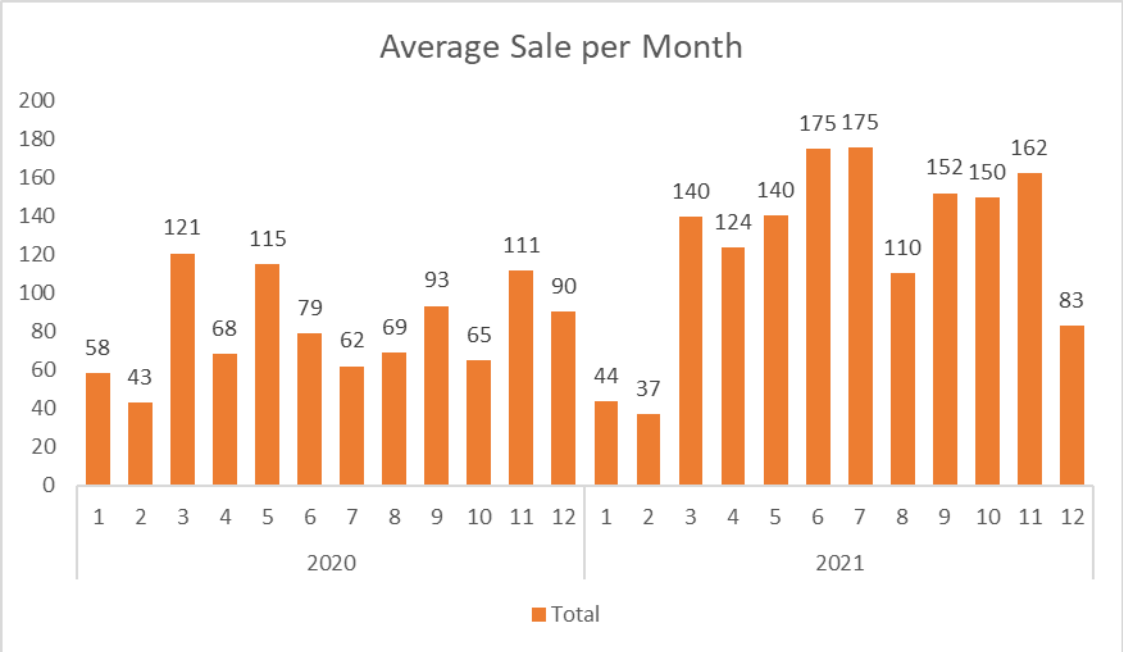


Chart 17 - Μέσος όρος καλαθιού ανά μήνα

Από το παραπάνω Bar Chart που απεικονίζει το average ποσό πώλησης ανά μήνα για τη διετία 2020-2021 προκύπτει ότι παρόλο που τους μήνες Φεβρουάριο και Νοέμβριο ο αριθμός παραγγελιών αυξάνεται κατακόρυφα, το μέσο ποσό πώλησης ανά παραγγελία είναι χαμηλό, ενώ τους θερινούς

μήνες διαπιστώνεται ότι ο μέσος όρος είναι μεγαλύτερος. Κατά συνέπεια, η επιχείρηση πρέπει να εκμεταλλευτεί τα αποτελέσματα του MBA και σε συνδυασμό με τον ήδη αυξημένο όγκο παραγγελιών να αυξήσει το μέσο κόστος του κάθε καλαθιού για τους παραπάνω μήνες.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

7

Τεχνικές λεπτομέρειες

7.1 Πλατφόρμες και προγραμματιστικά εργαλεία

Για την εκτέλεση των πειραμάτων στα πλαίσια της έρευνας χρησιμοποιήθηκαν οι εφαρμογές της SAS Institute και πιο συγκεκριμένα, η εφαρμογή SAS Enterprise Guide και η εφαρμογή SAS Enterprise Miner.

7.2 Λεπτομέρειες υλοποίησης

7.2.1 Μοντέλο RFM

Το Μοντέλο RFM από το οποίο προκύπτει το Clustering των πελατών παράγεται από τα Original Data. Τα μη επεξεργασμένα δεδομένα δεν προσφέρουν την απαραίτητη πληροφόρηση από μόνα τους. Συνεπώς, χρησιμοποιώντας και συνδυάζοντας τις υφιστάμενες πληροφορίες καταλήγουμε στο ιδανικό Dataset το οποίο θα χρησιμοποιηθεί στο SAS Enterprise Miner.

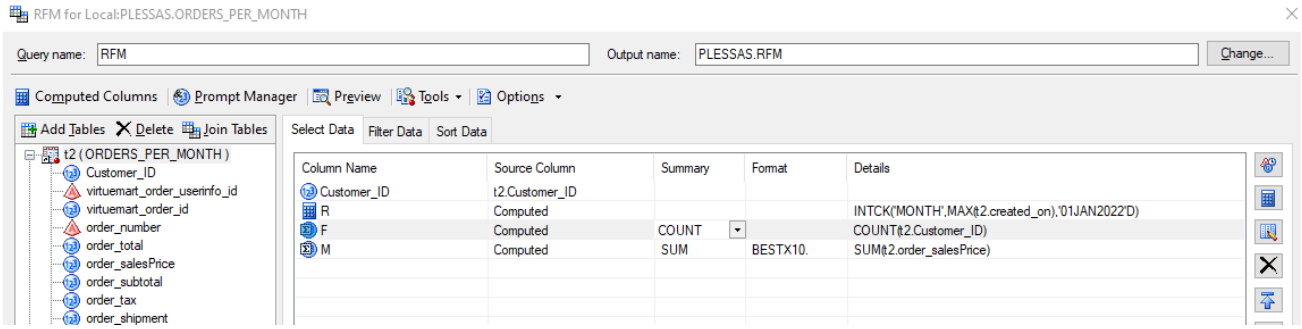
Αρχικά, για την παραγωγή του dataset πρέπει να συλλεχθούν οι κατάλληλες μεταβλητές και έτσι επιλέγονται οι παρακάτω:

- Customer ID
- Virtuemart_Order_id

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

- Order_salesPrice
- created_on

Για το R υπολογίζονται οι μήνες από την ημερομηνία 01/01/2022, ενώ για το F αθροίζονται όλες οι παραγγελίες ανά πελάτη. Τέλος, για το M αθροίζεται από όλες τις παραγγελίες το ποσό πληρωμής.



Εικόνα 5 - SAS Enterprise Guide - παραμετροποίηση query για την παραγωγή του RFM Dataset

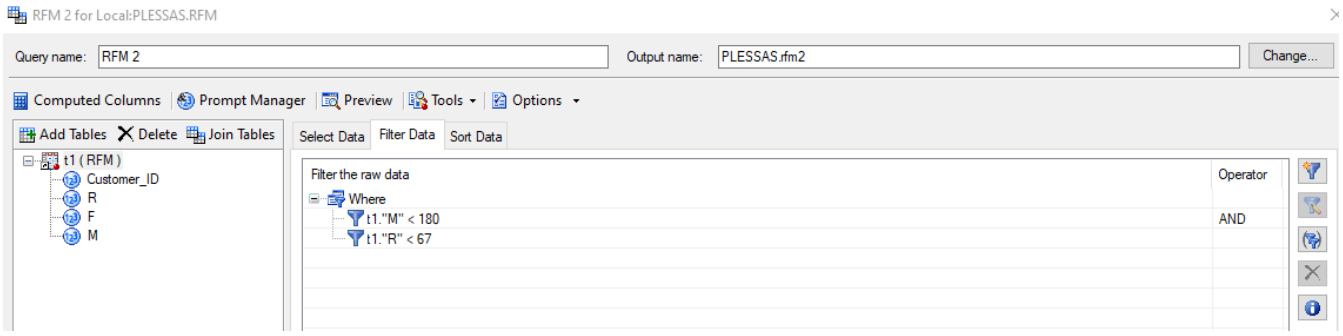
```
%_eg_conditional_dropds(PLESSAS.RFM);

PROC SQL;
   CREATE TABLE PLESSAS.RFM AS
   SELECT DISTINCT t2.Customer_ID,
          /* R */
            (INTCK('MONTH',MAX(t2.created_on),'01JAN2022'D)) LABEL="R" AS R,
          /* F */
            (COUNT(t2.Customer_ID)) LABEL="F" AS F,
          /* M */
            (SUM(t2.order_salesPrice)) FORMAT=BESTX10. LABEL="M" AS M
      FROM PLESSAS.ORDERS_PER_MONTH t2
      WHERE t2.created_on > '28Dec2019'd
      GROUP BY t2.Customer_ID;
QUIT;
```

Εικόνα 6 - Κώδικας SQL για το RFM Query

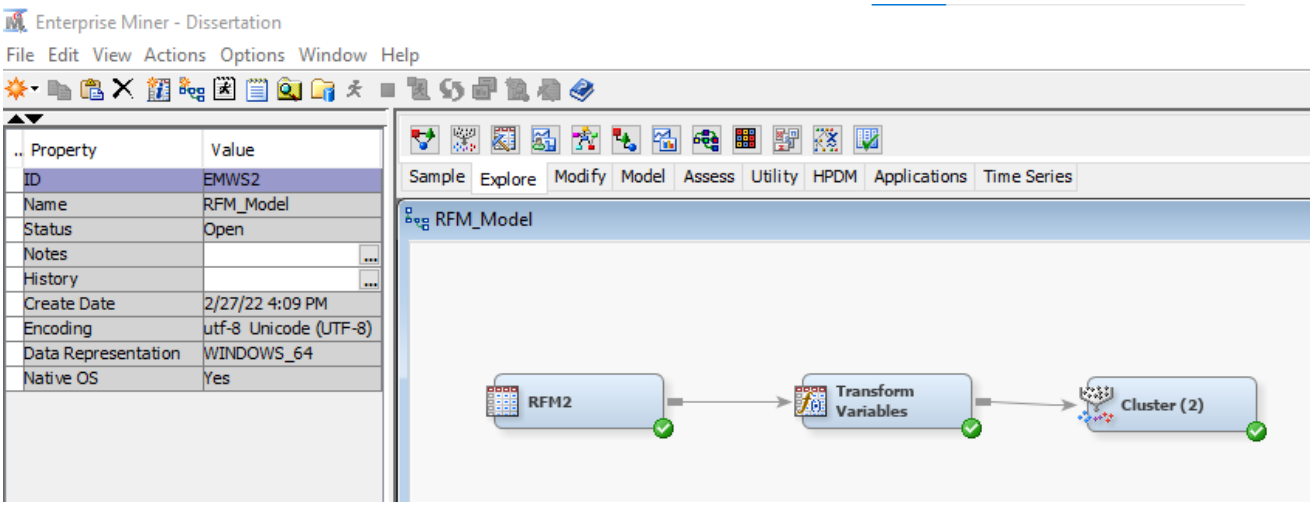
Μέσω της απεικόνισης των boxplots εντοπίζονται τα Outliers που πρέπει να αφαιρεθούν. Για την αφαίρεση των τιμών αυτών παραμετροποιουνται φίλτρα στο query του RFM και συνεπώς παράγεται το dataset RFM2.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ



Εικόνα 7 - SAS Enterprise Guide - Filtering ακραίες τιμές

Στη συνέχεια μεταφέρεται στο SAS Enterprise Miner με τη δημιουργία ροής διαγράμματος.

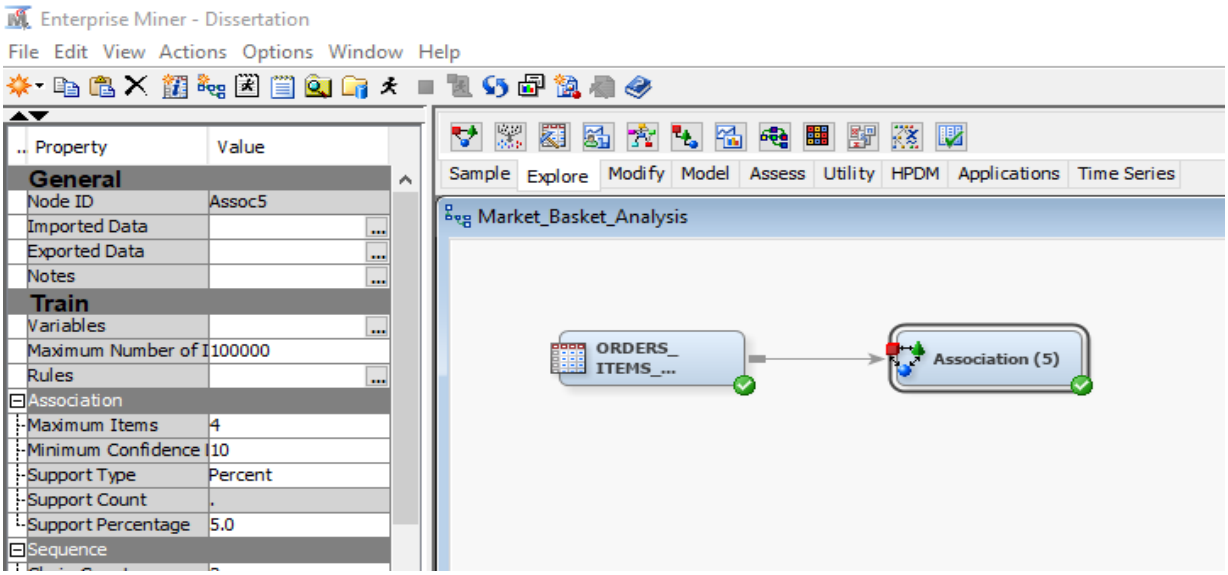


Εικόνα 8 - SAS Enterprise Miner - Diagram Flow RFM

7.2.2 Market Basket Analysis

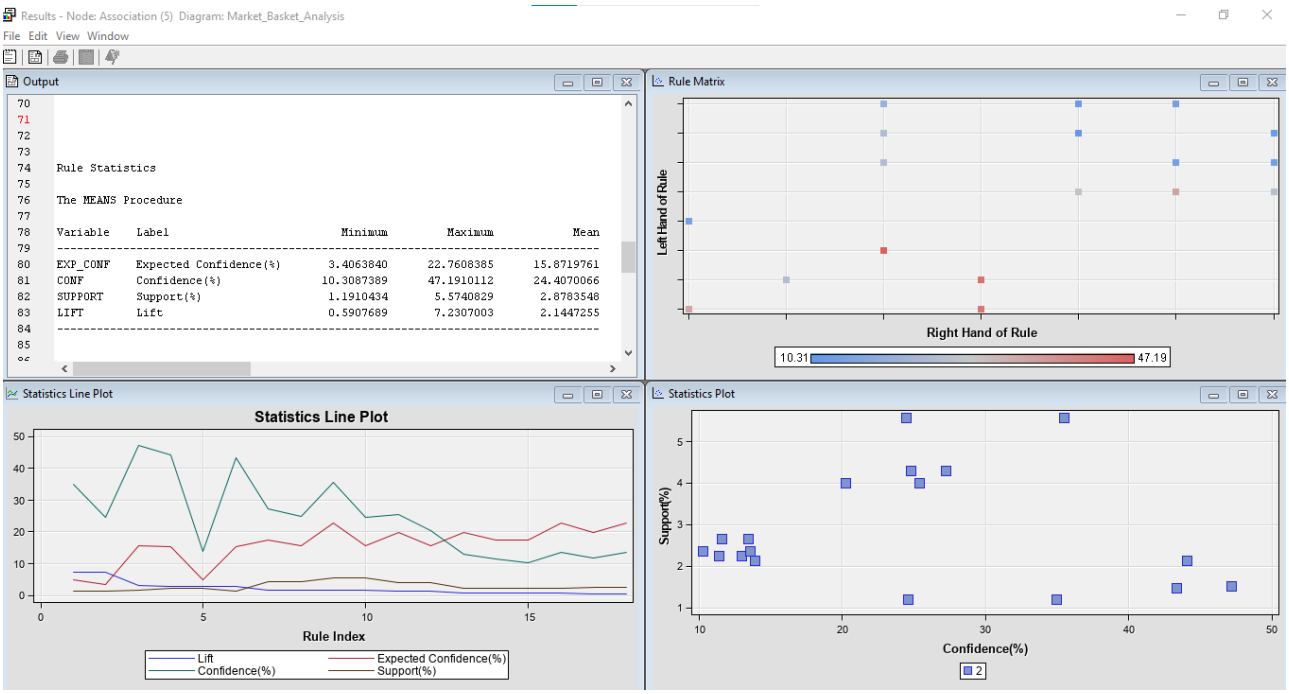
Για την εφαρμογή του μοντέλου MBA στο SAS Enterprise Miner χρησιμοποιήθηκε το αρχείο «Order Items». Το αρχείο, όπως παράγεται από την βάση δεδομένων του ηλεκτρονικού καταστήματος, εξυπηρετούσε το σκοπό της εφαρμογής του μοντέλου και δε χρειαζόταν περαιτέρω επεξεργασία. Οι εγγραφές είναι καταγεγραμμένες ανά SKU με επιπρόσθετες πληροφορίες τον αριθμό παραγγελίας, την περιγραφή και την κατηγορία του προϊόντος. Στο παρακάτω στιγμιότυπο από το λογισμικό διακρίνονται αριστερά η παραμετροποίηση του min Confidence 10% και του min Support 5%. Επειδή οι κανόνες που εξορύχθηκαν ήταν ικανοποιητικοί, δεν απαιτήθηκε περαιτέρω παραμετροποίηση. Εάν οι κανόνες ήταν υπερβολικά πολλοί, θα έπρεπε σταδιακά να αυξήσουμε το min Confidence και το min Support, ώστε να μειωθούν.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ



Εικόνα 9 - SAS Enterpise Miner - Diagram Flow Market Basket Analysis

Το λογισμικό επίσης παράγει στατιστικά στοιχεία που υποστηρίζουν τον ερευνητή στη διατύπωση των συμπερασμάτων του.



Εικόνα 10 - SAS Enterpise Miner - Statistics Charts

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

```
*------------------------------------------------------------*,
* Assoc5: Creating RULEMAP and Output data sets;
*------------------------------------------------------------*;
%let _scoreDs = &EM_SCORE_OUTPUT;
proc sort data=&_scoreDs;
by virtuemart_order_id;
run;
proc mbscore data=&_scoreDs out=score_ruleid INCLUDE ALL_ID
;
customer virtuemart_order_id;
target Category;
;
rules data=work.ruleid;
run;
data &_scoreDs;
set score_ruleid;
array _r{18} _r1-_r18 (18*0);
by virtuemart_order_id;
if first.virtuemart_order_id then do;
do i=1 to 18;
_r[i]=0;
end;
end;
```

Εικόνα 11 - Μέρος κώδικας από το λογισμικό

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

8

Επίλογος

Οι εφαρμογές Data mining Techniques είναι απαραίτητα εργαλεία στο χώρο του Marketing, από τα οποία εξάγονται χρήσιμα συμπεράσματα, είτε με πρόβλεψη, είτε με ανάλυση ιστορικών στοιχείων και διευκολύνουν τους εμπορικούς οργανισμούς, ώστε να λάβουν τις σωστές αποφάσεις. Η υιοθέτησή τους είναι αυξανόμενη, καθώς βοηθούν να απαντηθούν σημαντικά ερωτήματα με την επεξεργασία δεδομένων, τα οποία πλέον συγκεντρώνονται πολύ εύκολα. Επιπρόσθετα, το κόστος αποθήκευσης δεδομένων συνεχώς μειώνεται, κάτι το οποίο σημαίνει ότι οι επιχειρήσεις συλλέγουν πιο τακτικά και δομημένα τα δεδομένα τους. Πιο αναλυτικά οι τεχνικές RFM και Market Basket Analysis που αναλύθηκαν στη διπλωματική, είχαν ως στόχο την προσέγγιση της καταναλωτικής συμπεριφοράς των πελατών και κατ' επέκταση την εξατομικευμένη εφαρμογή προωθητικών ενεργειών. Στην Ελλάδα, όπως αναφέραμε, η αύξηση των ηλεκτρονικών καταστημάτων ήταν ραγδαία τα τελευταία χρόνια και προβλέπεται να συνεχίσει να αυξάνει. Στοιχεία από έκθεση της European Commission αναφέρουν ότι οι μικρές επιχειρήσεις στην Ελλάδα αποτελούν το 97% του συνόλου των επιχειρήσεων. Αντίστοιχα, ο μέσος όρος στην Ευρώπη είναι 93% (2020). Επειδή η φιλοσοφία της λειτουργίας μιας ατομικής επιχείρησης είναι σχεδόν παράδοση στην Ελλάδα, οι περισσότεροι ιδιοκτήτες ασκούν μόνοι τους στρατηγικές Marketing, ενώ το segmentation των πελάτων πραγματοποιείται με εμπειρικά στοιχεία. Κάτι τέτοιο οδηγεί πολλές φορές σε ολέθρια λάθη, τα οποία συνήθως απειλούν τη βιωσιμότητα του οργανισμού.

Στην υπό εξέταση επιχείρηση μέσω της εφαρμογής του RFM επιτεύχθηκε το Segmentation των πελατών σε 5 κατηγορίες. Τους Promising, τους New Customers, τους Lost, τους Need Attention και

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

τους High Value. Για κάθε ομάδα θα πρέπει να λάβει υπόψη της η επιχείρηση τα σημεία που επισημάνθηκαν σχετικά με τη συχνότητα παραγγελιών, την πολύμηνη απουσία τους και τέλος, το συνολικό ποσό που έχουν ξοδέψει σε αγορές οι πελάτες. Για κάθε Cluster πελατών μπορούν να εφαρμοστούν διαφορετικές στρατηγικές Marketing, όπως η υιοθέτηση ενός Loyalty program και η αποστολή προωθητικών ενεργειών ή εκπτωτικών κουπονιών μέσω Direct mailing.

Επιπλέον, με την εφαρμογή του Market Basket Analysis στις συναλλαγές της επιχείρησης προέκυψαν οι συσχετίσεις στις αγορές προϊόντων από τους πελάτες. Από τα αποτελέσματα αυτά αντλεί πληροφόρηση το τμήμα Marketing για τις σχέσεις που δημιουργούνται μεταξύ των κατηγοριών των προϊόντων. Τέτοιες κατηγορίες είναι τα χριστουγεννιάτικα αξεσουάρ με φωτάκια LED και οι Αποκριάτικες Περούκες με Αποκριάτικα Αξεσουάρ. Το MBA επιτρέπει έτσι στους ερευνητές να καταλήγουν σε βάσιμες υποθέσεις, χρησιμοποιεί δεδομένα που συχνά θεωρούνται ασήμαντα, ενώ δημιουργεί δυναμικές θεωρίες για την αξιολόγηση σχέσεων σε πολλαπλά επίπεδα.

Η εφαρμογή των τεχνικών δεν περιορίζονται μόνο στους παραπάνω σκοπούς. Οι δυνατότητες που προσφέρονται στους ερευνητές είναι μεγάλες. Είναι εφικτή η εφαρμογή των τεχνικών συνδυαστικά, όπως για παράδειγμα η εφαρμογή MBA για το Cluster 3 με τους Lost πελάτες, ώστε να διερευνηθούν οι προτιμήσεις τους και στοχευμένα να συσταθούν προωθητικές καμπάνιες για να επιστρέψουν όσο το δυνατόν περισσότεροι για αγορές. Ακόμα περισσότερο θα έχει ενδιαφέρον η ανάλυση των συναλλαγών και το Segmentation των πελατών του φυσικού καταστήματος, ώστε να συγκριθούν τα αποτελέσματα και να γίνουν περισσότερο κατανοητές οι συμπεριφορές των πελάτων ανάλογα με τον τρόπο που αγοράζουν. Τέλος, μελλοντικές έρευνες θα μπορούσαν να πραγματοποιηθούν ως επέκταση του MBA στην πρόβλεψη προμήθειας των προϊόντων για το retail κατάστημα. Για παράδειγμα, γνωρίζοντας πότε ξεκινά η περίοδος των Αποκριών και των Χριστουγέννων, να πραγματοποιούνται παραγγελίες με πρόβλεψη από την επιχείρηση, ώστε να καλύπτονται οι ανάγκες των πελατών χωρίς η επιχείρηση να τεθεί Out of stock σε ορισμένα προϊόντα.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

9

Bibliography

Aguinis, H., Forcum, L., & Joo, H. (2013). Using Market Basket Analysis in Management Research. *Journal of Management* , 1799-1824.

Alfarizi, M., & Sari, R. (2021). Analysis Of Factors Affecting Customer Behavior Of Marketplace Applications: A Case Study Of Cash On Delivery (COD) Payment Systems.

Baker

Bult, J., & Wansbeek, T. (1995). Optimal Selection for Direct Mail. 378-394.

Chen, D., Sain, S., & Guo, K. (2012). *Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining.* Macmillan Publishers Ltd.

Douglas, L. (2001). *3D Data Management: Controlling Data Volume, Velocity and Variety.*

Dunham, M. (2002). *Data Mining - Introductory and Advanced Topics.*

Forghani, E., Sheikh, R., Hosseini, S., & Sana, S. (2021). The impact of digital marketing strategies on customer's buying behavior in online shopping using the rough set theory.

Hiziroglu, A. (2013). Soft computing applications in customer segmentation: State-of-art review and critique. Στο *Expert Systems with Applications* (σσ. 6491-6507).

Jarboe, G. (2015, 03 16). Ανάκτηση από Tubular insights: https://tubularinsights.com/mcdonalds-our-food-your-questions-video-campaign/

J-C, F. (2005). *Problems, methods, and challenges.*

Johnson, J. (January 2022). *Global internet access rate 2005-2021.*

Kahan, R. (1998). Using database marketing techniques to enhance your one to-one marketing initiatives. *JOURNAL OF CONSUMER MARKETING*, 491-493.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Klein, S. (2017). *The world of big data and IoT.* IoT solutions in Microsoft's azure IoT suite. Springer, New York.

Kumar V., A. T. (2014). Role of ICT in driving e-commerce business in developing countries . *Department of Information Technology, BGSB University, Rajour (J&K)*, (σ. 28).

Kurniawana, F., Umayaha, B., Hammad, J., Susiki Nugroho, S. M., & Hariadi, M. (2018). Market Basket Analysis to Identify CustomerBehaviorsby Way of Transaction Data. 20-25.

Little, R. J., & Rubin, D. (2020). *Statistical Analysis with Missing Data.*

Müller, H., & Freytag, J.-C. (2005). *Problems, methods, and challenges.* Professoren des Inst. Für.

Russell, G., Ratneshwar, S., Shocker, A., Bell, D., Bodapati, A., Degeratu, A., . . . Shankar, V. (1999). Multiple-category decision-making: Review and synthesis. Marketing Letters. 319-332.

Sharanjit Kaur, S. (2021). *Customer Segmentation Using Clustering Algorithm.* International Conference on Technological Advancements and Innovations (ICTAI).

Sharma, R. (2020, 12 01). *https://www.upgrad.com/blog*. Retrieved from https://www.upgrad.com/blog: https://www.upgrad.com/blog/types-of-data/

Smith, W. (1956). Product differentiation and market segmentation as an alternative marketing strategy. *Journal of Marketing*, 3-8.

Sumit Koul, T. M. (2021). *Customer Segmentation Techniques on E-Commerce.* Institute of Electrical and Electronics Engineers.

Ting, P.-H., Pan, S., & Chou, S.-S. (2010). Finding Ideal Menu Items Assortments: An Empirical Application of Market Basket Analysis.

Wei, J.-T., Lin, S.-Y., & Wu, H.-H. (2010). A review of the application of RFM model. 4199-4206.

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Appendices

Λίστα πινάκων

Εικόνα 1 - 6Vs of Big Data 3
Εικόνα 2 - Data Cleansing Flow (Müller & Freytag, 2005) 16
Εικόνα 3 - Data Mining Models and Tasks Dunham, 2002) 17
Εικόνα 4 - RFM Flow 21
Εικόνα 5 - SAS Enterprise Guide - παραμετροποίηση query για την παραγωγή του RFM Dataset .. 59
Εικόνα 6 - Κώδικας SQL για το RFM Query 59
Εικόνα 7 - SAS Enterprise Guide - Filtering ακραίες τιμές 60
Εικόνα 8 - SAS Enterpise Miner - Diagram Flow RFM 60
Εικόνα 9 - SAS Enterpise Miner - Diagram Flow Market Basket Analysis 61
Εικόνα 10 - SAS Enterpise Miner - Statistics Charts 61
Εικόνα 11 - Μέρος κώδικας από το λογισμικό 62

Λίστα Charts

Chart 1 - Χρήση Internet 2005 -2021 (Johnson, January 2022) 4
Chart 2 - Internet users who bought or ordered goods or services for private use in the previous 12 months by age group, EU, 2011-2021 6
Chart 3 - Global retail ecommerce sales worldwide from 2014 to 2024 (in billion USD) 6
Chart 4 - Customers per Gender 38
Chart 5 - Customers Group of Ages 39
Chart 6 - Πελάτες ανά Διοικητική Περιφέρεια 40
Chart 7 - Πωλήσεις ανά έτος 41
Chart 8 - Πωλήσεις με βάση τον τρόπο πληρωμής 42
Chart 9 - Πωλήσεις ανά τρόπο πληρωμής ανά έτος 42

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Chart 10 - Box Plots for Recency, Frequency and Monetary 44
Chart 11 - Box Plots for Recency και Monetery μετά την αφαίρεση ακραίων τιμών 45
Chart - % πωλήσεων ανά μήνα μέσα στο έτος 49
Chart - % πωλήσεων ανά μήνα 2020-2021 50
Chart - Πωλήσεις ανά κατηγορία προϊόντος 51
Chart - Διάγραμμα παραγόμενων κανόνων 53
Chart - Διάγραμμα κανόνων μετά αύξηση Support & Confidence 54
Chart - Μέσος όρος καλαθιού ανά μήνα 56

ATHENS UNIVERSITY OF ECONOMICS & BUSINESS
ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ