



**ΑΡΙΣΤΟΤΕΛΕΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ
ΘΕΣΣΑΛΟΝΙΚΗΣ
ΣΧΟΛΗ ΘΕΤΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΑΝΑΛΥΤΙΚΗ ΕΠΙΧΕΙΡΗΜΑΤΙΚΟΤΗΤΑ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΜΕ ΤΗ ΧΡΗΣΗ ΤΟΥ POWER BI

ΣΠΑΝΟΥ ΑΛΕΞΑΝΔΡΑ

**ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ:
ΒΛΑΧΑΒΑΣ ΙΩΑΝΝΗΣ**

ΘΕΣΣΑΛΟΝΙΚΗ 2018

*Η παρούσα εργασία είναι
αφιερωμένη στον φίλο
μου τον Κώστα*

ΠΕΡΙΛΗΨΗ

Τα τελευταία χρόνια όλο και περισσότερες επιχειρήσεις επενδύουν σε εργαλεία τεχνολογίας με σκοπό την ανάλυση κι αξιολόγηση των επιχειρηματικών δεδομένων που διατηρούν. Ο τομέας της Επιχειρηματικής Ευφυΐας έχει επηρεάσει σημαντικά όχι μόνο τις διαδικασίες της ανάλυσης των δεδομένων, αλλά και τις διαδικασίες των λήψεων αποφάσεων. Η προσπάθεια μετατροπής των δεδομένων σε πολύτιμη πληροφορία και κέρδος έχουν καταστήσει τη δημιουργία των αναφορών απαραίτητη σε έναν οργανισμό. Τέτοιες αναφορές δημιουργούνται με τη χρήση των εργαλείων της Επιχειρηματικής Ευφυΐας και περιλαμβάνουν την παρουσίαση των αποτελεσμάτων μιας εξέτασης-ανάλυσης δεδομένων που αφορούν είτε το παρόν είτε το μέλλον.

Η παρούσα εργασία έχει ως αντικείμενο την παρουσίαση των τεχνολογιών που απαρτίζουν την Επιχειρηματική Ευφυΐα και των τεχνικών που χρησιμοποιεί. Ακόμη, πραγματοποιείται η δημιουργία αναφορών με τη χρήση του εργαλείου Power BI Desktop, εξετάζονται οι δυνατότητες που διαθέτει και πώς οι χρήστες μπορούν να τις αξιοποιήσουν ώστε να εξάγουν συμπεράσματα από τα επιχειρηματικά δεδομένα που εξετάζουν.

ABSTRACT

In recent years, more and more businesses have been investing in technology tools to analyze and evaluate their business data. Business Intelligence has significantly affected not only the processes of data analysis, but the decision-making processes too. The effort to convert the data into valuable information and profit has made the creation of reports essential to an organization. Such reports are generated using Business Intelligence tools and include the presentation of the results of a data analysis or examination that involves either the present or the future.

The particular work aims at presenting the technologies that constitute the Business Intelligence and the techniques it uses. In addition, it includes the creation of reports using Power BI Desktop tool, the examination of the capabilities it has and how users can use them in order to draw conclusions from the business data they analyze.

ΕΥΧΑΡΙΣΤΙΕΣ

Με την ολοκλήρωση της διπλωματικής μου εργασίας, αισθάνομαι την ανάγκη να εκφράσω την ευγνωμοσύνη μου σε όλους όσους με υποστήριξαν άμεσα ή έμμεσα στην εκπόνησή της. Αρχικά θα ήθελα να ευχαριστήσω θερμά τον καθηγητή και επιβλέποντα μου κ. Βλαχάβα Ιωάννη για την ανάθεση του θέματος, την άμεση ανταπόκρισή του στους προβληματισμούς μου, τις χρήσιμες επισημάνσεις του και το ενδιαφέρον του. Επίσης, επιθυμώ να ευχαριστήσω τον συνεπιβλέποντα της εργασίας μου Διδάκτορα κ. Φαχαντίδη Ανέστη για την συνεργασία του και την πολύτιμη βοήθειά του. Ακόμη, θα ήθελα να ευχαριστήσω όλους τους καθηγητές μου για τη γνώση που αποκόμισα και τους συμφοιτητές μου για τις συνεργασίες μας. Τέλος, θα ήθελα να ευχαριστήσω όλους τους ανθρώπους από το φιλικό μου περιβάλλον που πίστεψαν σ' εμένα και την οικογένειά μου για τη συνεχή υποστήριξή τους, η οποία ήταν καταλυτική στην πραγματοποίηση των μεταπτυχιακών σπουδών μου.

Φεβρουάριος 2018, Θεσσαλονίκη
Σπανού Αλεξάνδρα

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΕΡΙΛΗΨΗ.....	1
ABSTRACT.....	3
ΕΥΧΑΡΙΣΤΙΕΣ	5
ΠΕΡΙΕΧΟΜΕΝΑ.....	7
ΚΕΦΑΛΑΙΟ 1:	9
ΕΙΣΑΓΩΓΗ	9
ΚΕΦΑΛΑΙΟ 2:	11
ΕΠΙΧΕΙΡΗΜΑΤΙΚΗ ΕΥΦΥΙΑ	11
2.1 ΕΙΣΑΓΩΓΗ.....	11
2.2 ΕΙΔΗ ΕΠΙΧΕΙΡΗΜΑΤΙΚΗΣ ΕΥΦΥΙΑΣ.....	12
2.3 ΑΠΟΘΗΚΕΣ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΤΕΧΝΙΚΕΣ.....	13
2.3.1 ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΑΠΟΘΗΚΗΣ ΔΕΔΟΜΕΝΩΝ	14
2.3.2 ΔΙΑΔΙΚΑΣΙΕΣ ETL	15
2.3.3 ΣΧΕΔΙΑΣΜΟΣ ΑΠΟΘΗΚΗΣ ΔΕΔΟΜΕΝΩΝ.....	15
2.3.4 ΠΟΛΥΔΙΑΣΤΑΤΟ ΜΟΝΤΕΛΟ ΚΙ ΑΝΑΛΥΣΕΙΣ OLAP	17
2.3.5 ΠΡΟΫΠΟΘΕΣΕΙΣ ΜΙΑΣ ΑΠΟΘΗΚΗΣ ΔΕΔΟΜΕΝΩΝ.....	22
2.4 ΕΠΙΧΕΙΡΗΜΑΤΙΚΗ ΕΥΦΥΙΑ ΚΑΙ ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ.....	23
2.4.1 ΤΕΧΝΙΚΕΣ ΕΞΟΡΥΞΗΣ ΔΕΔΟΜΕΝΩΝ.....	25
2.4.2 ΔΕΝΤΡΑ ΑΠΟΦΑΣΗΣ	27
2.4.2.1 IDE Αλγόριθμος.....	28
2.4.3 ΝΕΥΡΩΝΙΚΑ ΔΙΚΤΥΑ	28
2.4.4 ΑΛΓΟΡΙΘΜΟΣ K-MEANS	29
2.4.5 ΑΛΓΟΡΙΘΜΟΣ APRIORI	30
2.5 ΠΛΕΟΝΕΚΤΗΜΑΤΑ-ΜΕΙΟΝΕΚΤΗΜΑΤΑ	32
2.5.1 ΠΛΕΟΝΕΚΤΗΜΑΤΑ ΕΠΙΧΕΙΡΗΜΑΤΙΚΗΣ ΕΥΦΥΪΑΣ.....	32
2.5.2 ΜΕΙΟΝΕΚΤΗΜΑΤΑ ΕΠΙΧΕΙΡΗΜΑΤΙΚΗΣ ΕΥΦΥΪΑΣ	33
2.6 Ο ΡΟΛΟΣ ΤΗΣ ΕΠΙΧΕΙΡΗΜΑΤΙΚΗΣ ΕΥΦΥΙΑΣ ΣΤΗ ΛΗΨΗ	
ΕΠΙΧΕΙΡΗΜΑΤΙΚΩΝ ΑΠΟΦΑΣΕΩΝ	34
2.7 ΣΥΣΤΗΜΑΤΑ ΕΠΙΧΕΙΡΗΜΑΤΙΚΗΣ ΕΥΦΥΙΑΣ ΚΑΙ ΠΛΗΡΟΦΟΡΙΑΚΑ	
ΣΥΣΤΗΜΑΤΑ (BI AND MIS)	36
ΚΕΦΑΛΑΙΟ 3:	39
POWER BI	39
3.1 ΕΙΣΑΓΩΓΗ.....	39
3.2 ΓΙΑΤΙ ΧΡΗΣΙΜΟΠΟΙΕΙΤΑΙ	39
3.3 ΤΟ ΠΕΡΙΒΑΛΛΟΝ ΤΟΥ POWER BI DESKTOP	40
ΚΕΦΑΛΑΙΟ 4:	53
ΜΕΛΕΤΗ ΠΕΡΙΠΤΩΣΗΣ ΜΕ ΤΗ ΧΡΗΣΗ ΤΟΥ POWER BI.....	53
4.1 ΠΕΡΙΓΡΑΦΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ	53

4.2 ΠΡΟ-ΕΠΕΞΕΡΓΑΣΙΑ ΤΩΝ ΔΕΔΟΜΕΝΩΝ.....	54
4.3 ΕΠΙΧΕΙΡΗΜΑΤΙΚΗ ΑΝΑΛΥΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΜΕ ΤΗ ΧΡΗΣΗ ΑΠΕΙΚΟΝΙΣΕΩΝ	55
4.3.1 ΣΥΣΤΑΔΟΠΟΙΗΣΗ ΜΕ ΤΟΝ ΚΩΔΙΚΟ ΤΟΥ ΠΡΟΪΟΝΤΟΣ.....	55
4.3.2 ΣΥΣΤΑΔΟΠΟΙΗΣΗ ΜΕ ΤΟΝ ΚΩΔΙΚΟ ΤΟΥ ΠΕΛΑΤΗ.....	61
4.3.3 ΔΗΜΙΟΥΡΓΙΑ ΑΝΑΦΟΡΩΝ ΚΑΙ ΧΡΗΣΗ ΣΥΣΤΑΔΩΝ	67
4.4 ΣΥΜΠΕΡΑΣΜΑΤΑ.....	75
ΚΕΦΑΛΑΙΟ 5:	77
ΣΥΜΠΕΡΑΣΜΑΤΑ	77
ΒΙΒΛΙΟΓΡΑΦΙΑ.....	81

ΚΕΦΑΛΑΙΟ 1:

ΕΙΣΑΓΩΓΗ

Στο σύγχρονο επιχειρησιακό κόσμο η αξία της πληροφορίας και η σωστή διαχείρισή της αποτελεί απαραίτητο παράγοντα όχι μόνο για την ορθή λειτουργία των διαδικασιών ενός οργανισμού, αλλά και για την εξέλιξη της πορείας του. Η άμεση πρόσβαση και η δυνατότητα παρακολούθησης των δεδομένων μέσα σε έναν οργανισμό αποτελεί θεμέλιο για την αξιολόγηση και την εκτίμηση της κατάστασης που διατρέχει. Οι ταχείς αλλαγές στον κόσμο του επιχειρείν έχουν μεγεθύνει τη σημασία του χρόνου στη διαδικασία της λήψης αποφάσεων και καθιστούν αναγκαία τη δυνατότητα πρόβλεψης πιθανών γεγονότων από τους οργανισμούς.

Τα επιχειρηματικά δεδομένα αποτελούν περιουσιακό στοιχείο μιας οργάνωσης και με τη συμβολή της Επιχειρηματικής Ευφυΐας μπορούν να αξιοποιηθούν αποτελεσματικότερα. Η αναζήτηση και η ανακάλυψη μιας σημαντικής πληροφορίας μέσα σε έναν οργανισμό μπορεί να οδηγήσει σε κέρδος ή μέχρι και σε κάποιο επιχειρηματικό άνοιγμα. Τα τελευταία χρόνια έχουν αναπτυχθεί αρκετές τεχνικές κι εργαλεία βασισμένα στην στατιστική και στην εξόρυξη δεδομένων με σκοπό την ανάλυση των επιχειρηματικών δεδομένων. Οι περισσότερες επιχειρήσεις στις μέρες μας έχουν στη διάθεσή τους μεγάλους όγκους δεδομένων, όπου με την κατάλληλη επεξεργασία τους μπορούν να αποκομίσουν πολύτιμη γνώση και συμπεράσματα για την πορεία και την εξέλιξή τους. Η δημιουργία αναφορών, δηλαδή η απεικόνιση των επιχειρηματικών δεδομένων εντός ή εκτός του οργανισμού αποτελεί απαραίτητη διαδικασία.

Αντικείμενο της παρούσας εργασίας είναι η ανάλυση και η δημιουργία αναφορών απεικόνισης των επιχειρηματικών δεδομένων με τη χρήση του Power BI, το οποίο διαθέτει εργαλεία επιχειρηματικής ανάλυσης.

Η εργασία δομείται σε κεφάλαια ως εξής:

- Στο Κεφάλαιο 2, γίνεται ανάλυση της έννοιας της Επιχειρηματικής Ευφυΐας και παρουσιάζονται τα μέρη που απαρτίζουν τη δομή της και οι τεχνικές που χρησιμοποιεί. Ακόμη, περιγράφονται οι μορφές σχεδιασμού μιας Αποθήκης Δεδομένων, η ερμηνεία του πολυδιάστατου μοντέλου και παρουσιάζονται μερικές τεχνικές που πραγματοποιούνται στην πολυδιάστατη ανάλυση. Επίσης, γίνεται ανάλυση των τεχνικών της Εξόρυξης Δεδομένων που χρησιμοποιεί η Επιχειρηματική Ευφυΐα κι αναφέρονται τα πλεονεκτήματα και τα μειονεκτήματα που προσδίδει σε έναν οργανισμό που επενδύει σε αυτήν.
- Στο Κεφάλαιο 3, δίνονται η έννοια και οι λόγοι της χρήσης του Power BI Desktop και παρουσιάζεται το περιβάλλον του. Ακόμη, αναλύονται οι δυνατότητες και τα εργαλεία που διαθέτει με την εφαρμογή ενός παραδείγματος στο πρόγραμμα. Επίσης, το κεφάλαιο αυτό περιλαμβάνει την παρουσίαση μερικών μορφών απεικονίσεων, που χρησιμοποιούνται στις δημιουργίες αναφορών, με βάση το παράδειγμά μας.
- Στο Κεφάλαιο 4, πραγματοποιείται μια μελέτη περίπτωσης στο πρόγραμμα του Power BI Desktop και χωρίζεται σε δύο μέρη. Στο πρώτο μέρος γίνεται η

παρουσίαση, η περιγραφή και η εξέταση των δεδομένων. Στο δεύτερο μέρος, τροφοδοτούνται τα δεδομένα μας στο πρόγραμμα, όπου εφαρμόζονται οι απαραίτητες τροποποιήσεις και χρησιμοποιώντας τα κατάλληλα εργαλεία που διαθέτει το πρόγραμμα γίνεται ανάλυση των επιχειρηματικών μας δεδομένων με τη δημιουργία αναφορών για διαφορετικές περιπτώσεις.

- Στο Κεφάλαιο 5, αναφέρονται η συμβολή των εργαλείων και των τεχνικών της Επιχειρηματικής Ευφυΐας στους οργανισμούς και τα συμπεράσματά μας για τη διαδικασία που ακολουθήσαμε κατά τη χρήση του εργαλείου του Power BI Desktop. Ακόμη, αναφέρεται η σημασία της δημιουργίας συστάδων-ομάδων και της χρήσης τους στις αναφορές σε έναν οργανισμό. Τέλος, κλείνουμε με κάποια σημαντικά σημεία των αναφορών, μερικούς προβληματισμούς κι ανοιχτά ζητήματα που θα μπορούσαν να διερευνηθούν από μια μελλοντική εργασία.
- Στο Παράρτημα I παρουσιάζονται αλφαβητικά η βιβλιογραφία και οι δικτυακοί τόποι που αναφέρονται στην εργασία.

ΚΕΦΑΛΑΙΟ 2:

ΕΠΙΧΕΙΡΗΜΑΤΙΚΗ ΕΥΦΥΙΑ

2.1 ΕΙΣΑΓΩΓΗ

Οι συνεχείς προκλήσεις στο επιχειρηματικό περιβάλλον έχουν δημιουργήσει την ανάγκη στους οργανισμούς να επενδύσουν στη χρησιμότητα και στη σημαντικότητα των πληροφοριών που παρέχουν, είτε στο εσωτερικό τους περιβάλλον είτε στο εξωτερικό. Η επιχειρηματική ευφυΐα παρέχει στους υπεύθυνους για τη λήψη αποφάσεων δεδομένα, πληροφορίες ή γνώσεις για την αντιμετώπιση των αποφάσεων που αφορούν συγκεκριμένα προβλήματα των επιμέρους φορέων λήψης αποφάσεων [1]. Οι εταιρίες αγωνίζονται ολοένα και περισσότερο με το μεγάλο όγκο των βάσεων δεδομένων και των εκτεταμένων αναλυτικών στοιχείων με στόχο τη βέλτιστη εκτίμηση κι αξιοποίηση αυτών στις αποφάσεις τους για την πραγματοποίηση των προσδοκιών τους.

Τα εργαλεία της επιχειρηματικής ευφυΐας είναι ευρέως αποδεκτά ως ένα νέο ενδιάμεσο λογισμικό μεταξύ των εφαρμογών συναλλαγών και των αιτημάτων υποστήριξης αποφάσεων [2], προσαρμόζοντας τη σύνδεση ή την αποσύνδεση των συστημάτων έχοντας ως σκοπό την αποτελεσματική υποστήριξη των επιχειρηματικών αποφάσεων. Οι δυνατότητες τους περιλαμβάνουν υποστήριξη αποφάσεων, OLAP, στατιστική ανάλυση, πρόβλεψη και εξόρυξη δεδομένων.

Οι Αποθήκες Δεδομένων (Data Warehouse) και οι τεχνικές OLAP (Online Analytical Processing) αποτέλεσαν σταθμό στην ιστορία της Επιχειρηματικής Ευφυΐας. Τα χαρακτηριστικά των αποθηκών δεδομένων, όπως η παροχή συστηματικής μεθόδου αποθήκευσης δεδομένων ως αρχείων, πινάκων ή αντικειμένων [3], η δυνατότητα ανάκτησης αυτών, ο περιορισμός πρόσβασης με γνώμονα την εξουσιοδότηση των υπεύθυνων χρηστών, η ασφάλεια και η ακεραιότητα [3] των δεδομένων και η παραμετροποίηση της δομής των δεδομένων αποτέλεσαν θεμέλιο για την λειτουργικότητα και την αποτελεσματικότητα της διαχείρισης και της επεξεργασίας μέσω των εργαλείων ανάλυσης που δημιουργήθηκαν μετέπειτα.

Οι τεχνικές OLAP οδηγούν τον υπεύθυνο στη λήψη αποφάσεων σε νέα επίπεδα στην ανάλυση δεδομένων. Με τη χρήση αυτών, η ανάλυση του υπεύθυνου λήψης αποφάσεων αλληλεπιδρά με τα δεδομένα που περιέχονται στο σύστημα [4]. Χρησιμοποιούν τα χρονικά χαρακτηριστικά που διατηρούνται στις αποθήκες δεδομένων, δίνοντας τη δυνατότητα πρόσβασης σε παρελθοντικά γεγονότα καθώς και μελλοντικά. Έτσι οι υπεύθυνοι έχουν το πλεονέκτημα εντοπισμού κάποιων τάσεων που κρύβονται πίσω από το πλήθος των δεδομένων και μπορούν να τις χρησιμοποιήσουν-εκμεταλλευτούν άμεσα για το παρόν, αλλά και για την πρόβλεψη μελλοντικών συνθηκών. Ακόμη, με την εξέταση των τάσεων αυτών μπορεί να προκύψουν συμπεράσματα πιθανών επιρροών σε έναν οργανισμό.

2.2 ΕΙΔΗ ΕΠΙΧΕΙΡΗΜΑΤΙΚΗΣ ΕΥΦΥΪΑΣ

Όπως κάθε τεχνολογικό εργαλείο έτσι κι η Επιχειρηματική Ευφυΐα έχει διάφορες μεθόδους αναλύσεων που επιλέγονται από τους αρμόδιους σύμφωνα με τους στόχους τους και κυρίως τα δεδομένα που έχουν στη διάθεσή τους. Πριν δούμε όμως τους τύπους αναλύσεων που υπάρχουν θα αναφέρουμε τα κύρια μέρη που απαρτίζουν την Επιχειρηματική Ευφυΐα, τα οποία είναι:

-Αποθήκες Δεδομένων

Οι Αποθήκες Δεδομένων αποτελούν τη βασική προϋπόθεση για την Επιχειρηματική Ευφυΐα, διότι σε αυτές διατηρείται όλο το υλικό των πληροφοριών που θα αντλήσει για να χρησιμοποιήσει. Μπορούν να περιέχουν κάθε είδους πληροφορία, είτε προέρχεται μέσα από τον οργανισμό είτε από το εξωτερικό του περιβάλλον. Παρέχουν τη δυνατότητα η δομή τους να καθοριστεί, δηλαδή να ελέγχεται η ροή των δεδομένων τους, και να τροποποιηθούν ώστε να τμηματοποιηθούν-οργανωθούν τα αρχεία τους.

-Πηγές Δεδομένων

Όπως οι τύποι των δεδομένων, έτσι κι οι πηγές τους μπορεί να είναι διαφορετικές, δηλαδή υπάρχει περίπτωση να είναι κάποιες βάσεις δεδομένων που δημιούργησαν τα μέλη ενός οργανισμού και τις διαχειρίζονται, οι οποίες συνήθως περιέχουν δομημένο υλικό ή μη, το οποίο μπορεί να προέρχεται από το παρελθόν μέσω διατήρησης ιστορικού δεδομένων ή από το διαδίκτυο ή από την εκτέλεση κάποιας δραστηριότητας που απαιτούσε αναζήτηση δεδομένων.

-Αγορά-Πεδία Δεδομένων

Κάθε στοιχείο μπορεί να κατηγοριοποιηθεί ανάλογα με το τμήμα που ανήκει ή με το πρόβλημα που μπορεί να επιλύσει. Η δημιουργία κάποιων πινάκων δεδομένων, που αφορούν συγκεκριμένα θέματα σε ένα τμήμα του οργανισμού βοηθά τους αρμόδιους, αλλά και τα εργαλεία της Επιχειρηματικής Ευφυΐας. Για παράδειγμα, το τμήμα των πωλήσεων έχει τα στοιχεία που το αφορούν. Ακόμη, θα μπορούσε να περιείχε πληροφορίες που αφορούν κάποιο θέμα αβεβαιότητας σύμφωνα με την άποψη των υπευθύνων.

-Εργαλεία Εξόρυξης Δεδομένων

Τα εργαλεία Εξόρυξης Δεδομένων δίνουν τη δυνατότητα στους χρήστες να ανακαλύπτουν γενικότητες, κανονικότητες και διάφορα πρότυπα και μοντέλα στις πηγές των δεδομένων. Ακόμη, με την εφαρμογή τους μπορούν οι χρήστες να κατανοήσουν τους κανόνες με τους οποίους προσαρμόζονται τα στοιχεία.

-Εργαλεία αιτημάτων-αναφορών

Οι τεχνικές και τα εργαλεία OLAP (Online Analytical Processing) χρησιμοποιούνται στις Αποθήκες Δεδομένων ή στα Πεδία Δεδομένων με σκοπό την ανάλυση των δεδομένων τους ώστε να αντλήσουν συμπεράσματα, όπως ανεύρεση τάσεων, ή απαντήσεις σε ερωτήματα που έχουν τα υπεύθυνα μέλη μιας οργάνωσης, όπως παράγοντες ενός συμβάντος. Επιπλέον, η δημιουργία αναφορών μπορεί να ενημερώσει τη διοίκηση κι ευρύτερα για την παρούσα κατάσταση του οργανισμού, αλλά και τη μελλοντική.

Όσον αφορά τους τύπους των αναλυτικών στοιχείων, είναι τρεις και παραθέτονται παρακάτω.

- Περιγραφικές Αναλύσεις
- Προγνωστικές Αναλύσεις
- Συντακτικές-Ρυθμιστικές Αναλύσεις

Οι Περιγραφικές Αναλύσεις παρέχουν πληροφορίες που αφορούν τον πραγματικό χρόνο ή τον παρελθοντικό. Χρησιμοποιούν την Εξόρυξη Δεδομένων και την Επιχειρηματική Ευφυΐα με στόχο να αντιμετωπίσουν λεπτομέρειες που αποκαλύπτουν συμπεράσματα για τις λειτουργικές δραστηριότητες ενός οργανισμού. Μπορούν να παρέχουν την εικόνα μιας επιχείρησης [5], όπως την κατάσταση της απόδοσής της ή των επιμέρους λειτουργιών της.

Οι Προγνωστικές Αναλύσεις χρησιμοποιούν στατιστικά μοντέλα και διάφορες τεχνικές σε τρέχοντα και παρελθοντικά στοιχεία με σκοπό τον οραματισμό συμβάντων. Για παράδειγμα, την πρόβλεψη των πωλήσεων με βάση τις προτιμήσεις και τη διαδικασία της αγοράς που πραγματοποιούν οι πελάτες ή την αξιολόγηση στα κοινωνικά μέσα [6].

Οι Συντακτικές Αναλύσεις δημιουργούν συγκεκριμένες συστάσεις δράσης για την επίτευξη ενός στόχου [7]. Αποτελούνται από μαθηματικές-στατιστικές μεθόδους που χρησιμοποιούνται με σκοπό να καθορίσουν τη βέλτιστη επιλογή ή λύση στα αιτήματα που υπάρχουν στους τομείς μιας επιχείρησης, όπως η καλύτερη στρατηγική τιμολόγησης [8]. Επιπλέον, δίνουν απαντήσεις σε υποθετικά ερωτήματα για το μέλλον, για παράδειγμα εάν κάποιος συνεργάτης ή προμηθευτής αποφασίσει να λήξει τη συνεργασία ή αν αλλάξει η παραγωγική διαδικασία ενός συγκεκριμένου προϊόντος.

2.3 ΑΠΟΘΗΚΕΣ ΔΕΔΟΜΕΝΩΝ ΚΑΙ ΤΕΧΝΙΚΕΣ

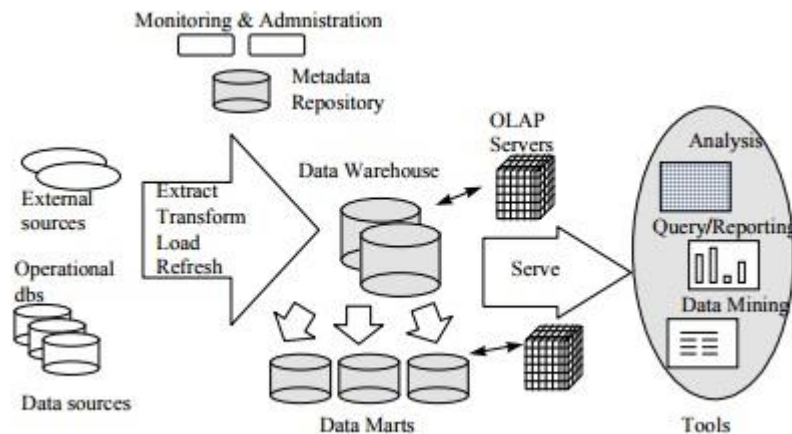
Ένα από τα σημαντικότερα περιουσιακά στοιχεία ενός οργανισμού είναι οι πληροφορίες που διατηρεί. Αυτό συνήθως διαχειρίζεται από έναν οργανισμό με δύο τρόπους, με τα λειτουργικά συστήματα καταγραφής και τις αποθήκες δεδομένων. Στην πραγματικότητα, η διαδρομή των δεδομένων είναι να εισαχθούν στα λειτουργικά συστήματα και ο ρόλος της αποθήκης δεδομένων είναι να υλοποιήσει την εξαγωγή τους. Οι χρήστες ενός λειτουργικού συστήματος ακολουθούν προκαθορισμένες διαδικασίες βασισμένες σε αυτές του οργανισμού. Αυτές οι διαδικασίες-δραστηριότητες αφορούν την λειτουργία του οργανισμού και οι χρήστες καθοδηγούνται από εντολές με σκοπό να διαχειρίζονται μια καταγραφή τη φορά. Για παράδειγμα, την εγγραφή νέου πελάτη ή την εκχώρηση κάποιων παραγγελιών.

Από την άλλη πλευρά, οι χρήστες μιας αποθήκης δεδομένων έχουν τον ρόλο της παρακολούθησης των προαναφερθέντων δραστηριοτήτων, δηλαδή μετρούν τις νέες παραγγελίες, συγκρίνουν παρελθοντικές μετρήσεις με τις νεότερες και προσπαθώντας να ερμηνεύσουν τα αποτελέσματα των συγκρίσεων ή μια σημαντική αλλαγή αναζητούν απαντήσεις θέτοντας ερωτήματα ανάμεσα σε μεγάλους όγκους δεδομένων. Η Αποθήκη Δεδομένων δηλαδή αποτελεί μια δομή δεδομένων που είναι εξελιγμένη σε θέματα που αφορούν τη διανομή, τη συνολική αποθήκευση, την πολύπλοκη διαχείριση και επεξεργασία ερωτημάτων [20]. Η λειτουργία της αφορά τη συλλογή εκτεταμένων όγκων δεδομένων από πολλαπλά λειτουργικά συστήματα, δεδομένα που μπορεί να πηγάζουν από το παρελθόν, την αποθήκευση αυτών και την

τροφοδότηση ενός ή περισσότερων πεδίων δεδομένων (Data Marts). Όπως έχουμε αναφέρει τα πεδία δεδομένων περιλαμβάνουν συγκεκριμένα θέματα ενός οργανισμού και μπορεί να αφορούν συγκεκριμένο τμήμα, όπως το τμήμα πωλήσεων. Οπότε, αντιλαμβανόμαστε το μέγεθος της σημαντικότητας και της αξίας των διαδικασιών που πραγματοποιεί μια αποθήκη δεδομένων.

2.3.1 Αρχιτεκτονική Αποθήκης Δεδομένων

Μία τυπική αρχιτεκτονική Αποθήκης Δεδομένων όπως φαίνεται στην Εικόνα 2.1 απαρτίζεται από τα εξής μέρη:



Εικόνα 2.1: Αρχιτεκτονική Αποθήκης Δεδομένων

(Πηγή: Chaudhuri & Dayal ,1997)

- Τα εργαλεία λογισμικού, τα οποία εξάγουν δεδομένα από πολλαπλές βάσεις δεδομένων, από διαφορετικές πηγές και του εξωτερικού περιβάλλοντος κι αναλαμβάνουν την προ-επεξεργασία αυτών των δεδομένων, δηλαδή την μορφοποίηση-παραμετροποίησή τους μέχρι την τοποθέτηση και την περιοδική ανανέωσή τους στην Αποθήκη Δεδομένων.

- Την κύρια Αποθήκη Δεδομένων.

- Τα Πεδία Δεδομένων, δηλαδή η ύπαρξη πολλών πινάκων δεδομένων με καθορισμένα θέματα, είτε εξωτερικά είτε ως υποσύνολα τμήματος της Αποθήκης Δεδομένων, όπου διαχειρίζονται και διατηρούνται από τεχνικές OLAP (Online Analytical Processing).

- OLAP servers, οι οποίοι διαθέτουν εργαλεία που υποστηρίζουν την αναζήτηση, την οπτικοποίηση και τη δημιουργία αναφορών των δεδομένων και εργαλεία Εξόρυξης Δεδομένων.

- Αποθήκη για διατήρηση και διαχείριση των μεταδεδομένων.

- Εργαλεία παρακολούθησης και διαχείρισης του συστήματος Αποθήκης Δεδομένων. (Analysis, Query Reporting, Data Mining)

Από τα βασικά μέρη αντιλαμβανόμαστε ότι οι Αποθήκες Δεδομένων αποτελούν μια αποθήκη πληροφοριών κατά την οποία συλλέγονται δεδομένα από διαφορετικές πηγές και δομές με τη χρήση κάποιων εργαλείων και αποτελούν βάση για τη μετέπειτα επεξεργασία με στόχο την εξαγωγή συμπερασμάτων και την ευρεία υποστήριξη μιας οργάνωσης.

2.3.2 Διαδικασίες ETL

Όπως έχει αναφερθεί ο όγκος των δεδομένων που διατηρεί μια Αποθήκη Δεδομένων προέρχεται από άλλες πηγές και δεν παράγεται από την ίδια. Στο αρχικό στάδιο με τη χρήση εργαλείων πραγματοποιείται η συλλογή, η μεταφορά, η παραμετροποίηση και η μεταφορά των δεδομένων. Οι διαδικασίες που αναλαμβάνουν αυτή τη διαχείριση των δεδομένων λέγονται εργασίες Εξαγωγής, Μετασχηματισμού και Φόρτωσης (Extract, Transform, Load (ETL)). Σύμφωνα με τα βήματα που ακολουθούν οι διαδικασίες αυτές μπορούν να χωριστούν σε ενότητες:

- Εξαγωγή Δεδομένων
- Επικύρωση Δεδομένων
- Καθαρισμός Δεδομένων
- Μετατροπή Δεδομένων
- Φόρτωση Δεδομένων

Όσον αφορά την εξαγωγή δεδομένων αποτελεί μια απαιτητική διαδικασία. Αρχικά θα πρέπει να υπάρχει γνώση της πηγής που προέρχονται τα δεδομένα, όπου μπορεί να είναι κάποιο επιχειρηματικό σύστημα ή κάποια βάση δεδομένων με προκαθορισμένα χαρακτηριστικά. Ακόμη, ο τρόπος που είναι δομημένα ή μη τα δεδομένα σε οποιαδήποτε πηγή αποτελεί σημαντικό γνώρισμα για τα επόμενα βήματα. Ο καθορισμός της δομής των δεδομένων περιλαμβάνεται σε αυτό το βήμα. Στη συνέχεια, κατά την επικύρωση των δεδομένων πραγματοποιούνται έλεγχοι που αφορούν την ποιότητα και την επάρκεια των δεδομένων, αφού στο σημείο αυτό είναι δυνατή η απόρριψη ή η δέσμευση για περεταίρω εξέταση.

Τα επόμενα βήματα που είναι ο καθαρισμός και η μετατροπή των δεδομένων συνήθως παρερμηνεύονται, διότι λανθασμένα θεωρούνται το ίδιο. Κατά τον καθαρισμό των δεδομένων καθήκον αποτελεί το φιλτράρισμα των ανεπιθύμητων δεδομένων. Τέτοια δεδομένα είναι τα ελλιπή, τα διπλότυπα και τα εσφαλμένα. Ενώ, η μετατροπή των δεδομένων αφορά την επεξεργασία που είναι εφικτό να υλοποιηθεί, όπως πράξεις πεδίου, επεξεργασία συμβολοσειρών χαρακτήρων, καθορισμός μηδενικής τιμής και άλλα.

Τέλος, κατά το τελευταίο βήμα των ETL διαδικασιών, γίνεται η μεταφορά των δεδομένων στον τελικό προορισμό, δηλαδή στην Αποθήκη Δεδομένων. Η βέλτιστη υλοποίηση της μεταφοράς των δεδομένων εξαρτάται από τον τύπο της λειτουργίας κι από τον όγκο των δεδομένων και μπορεί να γίνει με δύο τρόπους, με SQL ή προγράμματα συνολικής φόρτωσης.

2.3.3 Σχεδιασμός Αποθήκης Δεδομένων

Όπως αναφέρθηκε και παραπάνω μια Αποθήκη Δεδομένων είναι ένα σύστημα που ανακτά τα δεδομένα από διάφορες πηγές και τα τοποθετεί σε μια δομή, η οποία αποτελεί έναν αποθηκευτικό χώρο δεδομένων διαστάσεων ή έναν κανονικοποιημένο χώρο αποθήκευσης. Μερικές Αποθήκες Δεδομένων είναι σε μορφή διαστάσεων, αλλά ορισμένες είναι σε κανονικοποιημένη μορφή. Ο τρόπος που θα τοποθετηθούν τα δεδομένα σε έναν αποθηκευτικό χώρο αποτελεί σημαντικό παράγοντα για τη μετέπειτα διαχείριση των δεδομένων αυτών.

Μία βάση δεδομένων αποτελεί μια συλλογή δεδομένων, τα οποία έχουν σχέση μεταξύ τους και είναι δομημένα και αποθηκευμένα με κατάλληλο τρόπο. Το σχεσιακό μοντέλο δεδομένων (relational data model) αναπτύχθηκε από τον Codd (1970) και για

την αναπαράσταση των δεδομένων χρησιμοποιεί πίνακες, οι οποίοι έχουν διαφορετικό όνομα και αποτελούνται από ένα σύνολο γραμμών και στηλών. Κάθε γραμμή του πίνακα αποτελεί μια εγγραφή δεδομένων, ενώ οι στήλες περιλαμβάνουν τα χαρακτηριστικά της κάθε εγγραφής. Ένα παράδειγμα τέτοιου πίνακα βλέπουμε στον Πίνακα 2.1.

Κωδικός	Όνομα	Τηλέφωνο	Διεύθυνση
1985	Γιάννης	6923698	Δελφών 56
2632	Μαρία	6954264	Ακρίτας 5
7896	Αλεξία	69025687	Σεφέρη 6

Πίνακας 2.1: Παράδειγμα πίνακα

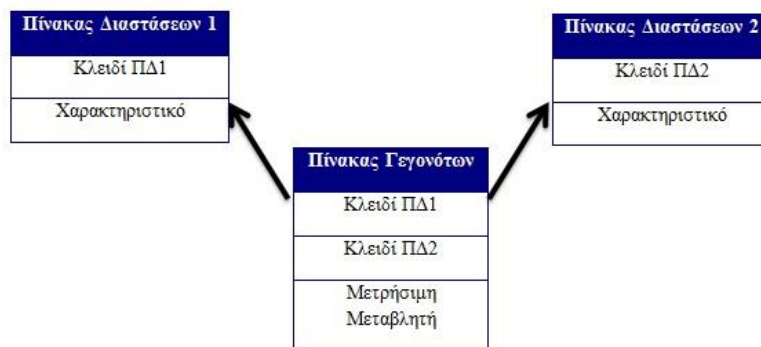
Μια βάση Δεδομένων μπορεί να περιλαμβάνει πολλούς τέτοιους πίνακες με διαφορετικά στοιχεία μεταξύ τους και διαφορετικό όγκο δεδομένων. Το χαρακτηριστικό που δημιουργεί μοναδικότητα στις γραμμές λέγεται πρωτεύον κλειδί. Η σχέση μεταξύ των διάφορων πινάκων που υπάρχουν μέσα σε μία βάση δεδομένων καθορίζεται από την ύπαρξη των πρωτεύοντων κλειδίων στους άλλους πίνακες, δηλαδή με βάση τον Πίνακα 2.1 να υπάρχει κάποιος άλλος πίνακας που θα περιλαμβάνει τη στήλη Κωδικός. Αυτό το σχεσιακό μοντέλο όμως καταγράφει λεπτομερώς τις εγγραφές και κατά τη δημιουργία σχέσεων μεταξύ των πινάκων δεσμεύεται μεγάλος όγκο. Η κανονικοποιημένη μορφή που αναφέρθηκε παραπάνω αποτελεί τη λύση για τη μείωση του δεσμευμένου χώρου, διότι έχουν ελαχιστοποιηθεί τα μεγέθη των δεδομένων που συνδέουν τους πίνακες μεταξύ τους.

Μια Αποθήκη Δεδομένων μπορεί να υλοποιηθεί με διαφορετικές μορφές σχημάτων όπως Σχήμα Αστέρα, Σχήμα Χιονονιφάδας. Ο λόγος που δημιουργήθηκαν αυτές οι μορφές είναι για να είναι δομημένοι οι πίνακες με πιο καθορισμένο τρόπο ώστε να επιτυγχάνονται οι εξαγωγές των δεδομένων που χρειάζονται για ανάλυση. Να υπάρχουν δηλαδή συνδέσεις μεταξύ των πινάκων που θα διευκολύνουν την αναζήτηση, επεξεργασία και εξαγωγή των δεδομένων.

Τα βασικά μέρη ενός μοντέλου διαστάσεων είναι οι πίνακες γεγονότων, οι οποίοι περιλαμβάνουν μετρήσιμα δεδομένα της επιχείρησης και οι πίνακες διαστάσεων, οι οποίοι περιέχουν χαρακτηριστικά κυρίως περιγραφικά για τον προσδιορισμό και την τμηματοποίηση των δεδομένων. Με αυτόν τον τρόπο δόμησης, η αναζήτηση, η περιήγηση κι ιδιαίτερα η ανάλυση απλοποιείται και είναι εφικτή μια πιο στοχευόμενη έρευνα των δεδομένων.

Σχήμα Αστέρα

Σε μια απεικόνιση σχήματος Αστέρα (Star Schema), μια διάσταση δεν θα είχε υποκατηγορίες, δηλαδή θα είχαμε έναν συνολικό πίνακα συμβάντων και γύρω του πίνακες διαστάσεων μόνο με τα χαρακτηριστικά τους δίχως περαιτέρω κατηγοριοποίησή. Δηλαδή, ο πίνακας συμβάντων θα περιείχε μετρήσιμες μεταβλητές εντός της επιχείρησης, όπως την ποσότητα ενός προϊόντος, όπου θα θέλαμε να εξετάσουμε-αναλύσουμε, και όλα τα πρωτεύοντα κλειδιά των πινάκων διαστάσεων που υπάρχουν γύρω του και σχηματίζουν την εικόνα ενός αστεριού. Ένα πρωτεύον κλειδί θα ήταν το κλειδί του πίνακα των προϊόντων, όπου θα περιείχε στοιχεία για τα προϊόντα. Ένα παράδειγμα σχήματος Αστέρα με δύο πίνακες διαστάσεων παρουσιάζεται στην Εικόνα 2.2.



Εικόνα 2.2: Σχήμα Αστέρα

Σχήμα Χιονονιφάδας

Σε μια απεικόνιση σχήματος Χιονονιφάδας (Snowflake Schema), μια διάσταση μπορεί να έχει υπό-κατηγορίες, δηλαδή θα είχαμε έναν συνολικό πίνακα γεγονότων και γύρω του πίνακες διαστάσεων όπου κάποιο από τα χαρακτηριστικά τους, για παράδειγμα η κατηγορία προϊόντος, θα μπορούσε να επεκταθεί σε άλλες υπο-ομάδες, όπως προϊόντα α και β τεχνολογίας. Ένα παράδειγμα σχήματος Χιονονιφάδας με δύο πίνακες διαστάσεων και με ένα χαρακτηριστικό που χωρίζεται παρουσιάζεται στην Εικόνα 2.3.



Εικόνα 2.3: Σχήμα Χιονονιφάδας

Η υλοποίηση μιας Αποθήκης Δεδομένων με τη μορφή του σχήματος αστέρα έχει το πλεονέκτημα ότι είναι πιο απλή σε σύγκριση με το σχήμα των χιονονιφάδων, που σημαίνει ότι οι διαδικασίες εισαγωγής και μεταφοράς των δεδομένων πραγματοποιούνται ευκολότερα. Βέβαια, υπάρχουν περιπτώσεις, όπως η εφαρμογή μιας ανάλυσης δεδομένων, όπου το σχήμα της χιονονιφάδας θα απέδιδε καλύτερα αποτελέσματα σε σύγκριση με το σχήμα αστέρα. Τέλος, με το σχήμα της χιονονιφάδας πραγματοποιείται τέτοιος διαχωρισμός όπου μειώνεται η πιθανή πλεονάζουσα χρήση χώρου σε μια δομή δεδομένων.

2.3.4 Πολυδιάστατο Μοντέλο κι Αναλύσεις OLAP

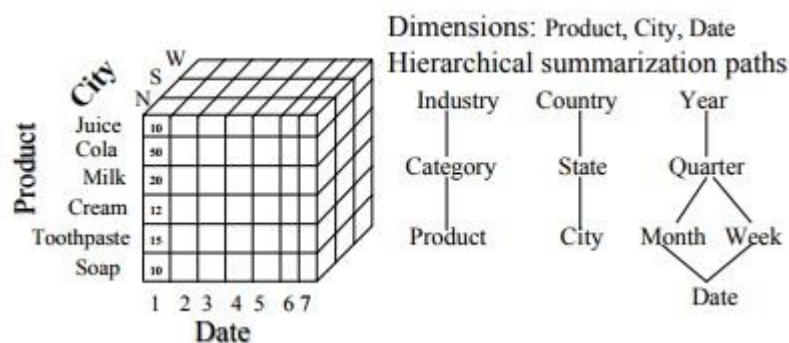
Οι περισσότερες επιχειρήσεις χρησιμοποιούν βάσεις δεδομένων τύπου OLTP (On Line Transaction Processing), όπου αυτός ο τύπος βάσεων δεδομένων περιέχει τις περισσότερες φορές πολλούς πίνακες, οι οποίοι συνήθως περιέχουν σχέσεις μεταξύ τους και πολλαπλές εγγραφές που μπορούν να επεκταθούν με προσθήκες νέων, να διαγραφούν και να ενημερώνονται συστηματικά. Παρά το γεγονός ότι οι OLTP

βάσεις δεδομένων φέρουν καλά αποτελέσματα στη συλλογή και τη διαχείριση των δεδομένων, πολλές φορές δεν αποδίδουν αποτελεσματικά σαν πηγές δεδομένων για τη δημιουργία αναφορών (reports). Οι δυνατότητές τους και η απόδοσή τους περιορίζονται σύμφωνα με κάποιους παράγοντες.

Αρχικά, η πολυπλοκότητα των δεδομένων αποτελεί κύρια αιτία, διότι όταν υπάρχει μεγάλος όγκος δεδομένων σε μια βάση δεδομένων OLTP προκύπτει μεγάλη δυσκολία στην κατανόηση των πινάκων και των σχέσεων που τους συνδέει μεταξύ τους ώστε να επιλεχθούν για τη συμμετοχή τους στην αναφορά. Ακόμη, η ύπαρξη πολλών μεμονωμένων αρχείων θα αποτελούσε πρόβλημα στη δημιουργία συγκεντρωτικών υπολογισμών και ομαδοποιημένων απεικονίσεων, όπου χρειάζονται ταξινόμηση των αρχείων σε εξέλιξη. Στην περίπτωση μεγάλου όγκου αρχείων θα μπορούσε να καθυστερήσει σημαντικά πολύ. Επιπλέον, η συχνότητα των μεταβολών των αρχείων που πραγματοποιούνται στις βάσεις δεδομένων θα οδηγήσουν σε ασυνεπή αποτελέσματα ανά τακτά χρονικά διαστήματα, διότι για παράδειγμα τα αποτελέσματα από τη δημιουργία μιας αναφοράς στηριζόμενη στα σημερινά δεδομένα δεν αποτελούν απαραίτητα ορθά κριτήρια για το μετέπειτα χρονικό διάστημα.

Μερικοί οργανισμοί για να αποφύγουν αυτούς τους περιορισμούς δημιουργούν τις αναφορές τους χρησιμοποιώντας τις βάσεις δεδομένων και τις τεχνικές OLAP (On Line Analytical Processing). Όπως αναφέρθηκε στην εισαγωγή της παρούσας εργασίας, οι τεχνικές OLAP χρησιμοποιούν τα χρονικά χαρακτηριστικά που διατηρούνται στις αποθήκες δεδομένων, δίνοντας τη δυνατότητα πρόσβασης σε παρελθοντικά γεγονότα καθώς και μελλοντικά. Τα δεδομένα της αποθήκης συνήθως διαμορφώνονται πολυδιάστατα, δηλαδή η δομή τους υποστηρίζεται από το πολυδιάστατο μοντέλο δεδομένων.

Ως πολυδιάστατο μοντέλο δεδομένων ή αναφέρεται και ως "κύβος δεδομένων" αποτελεί η πολυδιάστατη απεικόνιση δεδομένων, των οποίων η δομή είναι σχεδιασμένη με τον ίδιο τρόπο, δηλαδή πολυδιάστατη. Δηλαδή, στον κύβο δεδομένων η αποθήκευση των δεδομένων δεν είναι γραμμές και στήλες, όπως στο σχεσιακό μοντέλο δεδομένων, αλλά αποτελείται από γεγονότα και διαστάσεις [24]. Παραπάνω αναφερθήκαμε σε πίνακες συμβάντων και διαστάσεων. Φανταστείτε πως ένας κύβος δεδομένων αποτελείται από πολλούς τέτοιους πίνακες που έχουν ενωθεί και τον σχηματίζουν. Ένα παράδειγμα πολυδιάστατου μοντέλου παρουσιάζεται στην Εικόνα 2.4.



Εικόνα 2.4: Πολυδιάστατο Μοντέλο

(Πηγή: Chaudhuri & Dayal ,1997)

Για να γίνει πιο κατανοητό το πολυδιάστατο, κάθε τετραγωνάκι του πίνακα αποτελεί μια ή περισσότερες τιμές που λέγονται μέτρα. Όπως και στους πίνακες διαστάσεων

έτσι κι εδώ μια διάσταση μπορεί να έχει πολλά χαρακτηριστικά, τα οποία προσδιορίζουν μια επιχειρηματική δραστηριότητα, δηλαδή ένα γεγονός. Ο όγκος της ανάκτησης των πληροφοριών από έναν κύβο εξαρτάται από τις διαστάσεις που περιλαμβάνει, δηλαδή τα χαρακτηριστικά των διαστάσεων που αναλογούν στο γεγονός-συμβάν που θέλουμε να αναλύσουμε. Μπορούμε να αντιληφθούμε δηλαδή πως σύμφωνα με αυτό το μοντέλο μπορούμε να εξετάσουμε-αναλύσουμε και να απεικονίσουμε ένα γεγονός από πολλές διαστάσεις παράλληλα. Βέβαια, το μέγεθος της λεπτομέρειας που είναι δυνατό να αντλήσουμε σχετίζεται άμεσα με τη λεπτομέρεια που έχουν δομηθεί τα δεδομένα στην αποθήκη δεδομένων.

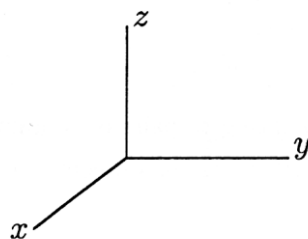
Όσον αφορά τη δομή των δεδομένων, σε μια βάση δεδομένων OLAP, όλες οι σχέσεις μεταξύ των δεδομένων έχουν προκαθοριστεί και αποθηκευτεί σε κύβους, οι οποίοι περιέχουν ιεραρχικές δομές, οι οποίες διαθέτουν εύκολη πλοήγηση στις διαστάσεις και στα μέτρα. Οπότε δεν είναι απαραίτητη η δημιουργία συνδέσεων ή η αναζήτηση της σύνδεσης κάποιου πίνακα δεδομένων με κάποιον άλλον. Επιπλέον, τα δεδομένα είναι εξίσου οργανωμένα, δηλαδή είναι ομαδοποιημένα, ταξινομημένα και έχουν συγκεντρωθεί στις βάσεις δεδομένων OLAP, οι οποίες περιλαμβάνουν μόνο στιγμιότυπά τους μιας και συνήθως είναι ιστορικά δεδομένα όπου και χρησιμοποιούνται αποκλειστικά για τη δημιουργία αναφορών. Στην περίπτωση προσθήκης νέων δεδομένων πραγματοποιείται η προσαρμογή τους στη βάση, αλλά τα δεδομένα που προϋπήρχαν δεν υφίστανται καμία αλλαγή, για αυτό και είναι εύκολη και συνεπής η ανάκτησή τους στις δημιουργίες απεικονίσεων.

Για την πολυδιάστατη οπτικοποίηση των δεδομένων χρησιμοποιούνται οι τεχνικές OLAP, κάποιες από τις οποίες είναι η Συναθροιστική Άνοδος (Roll-up), Αναλυτική Κάθοδος (Drill-Down), Οριζόντιος Τεμαχισμός (Slicing) και Κάθετος Τεμαχισμός (Dicing) [27].

• Συναθροιστική Άνοδος (Roll-up)

Με τη λειτουργία της συναθροιστικής ανόδου πραγματοποιείται η συσσωμάτωση σε έναν κύβο δεδομένων. Αυτό επιτυγχάνεται είτε αυξάνοντας το επίπεδο μιας ιεραρχίας για μια διάσταση είτε μειώνοντάς το προς τα κάτω. Ας δούμε ένα παράδειγμα:

Έστω ότι έχουμε τις εξής διαστάσεις, οι οποίες είναι ο Χρόνος σε εβδομάδες, οι Πωλήσεις, η Κατηγορία Προϊόντων και η Υποκατηγορία Προϊόντων. Φανταστείτε δηλαδή ένα γράφημα τριών διαστάσεων, όπως φαίνεται στην Εικόνα 2.5, όπου στον άξονα y είναι ο Χρόνος, δηλαδή οι εβδομάδες, στον άξονα z είναι η Κατηγορία των Προϊόντων και στον x η Υποκατηγορία των Προϊόντων.



Εικόνα 2.5: Γράφημα τριών διαστάσεων

Ένας τρόπος για να κάνουμε συναθροιστική άνοδο είναι να μειώσουμε το επίπεδο της ιεραρχίας προς τα κάτω. Αυτό μπορεί να επιτευχθεί ως εξής στο παράδειγμά μας. Έστω ότι έχουμε ένα κομμάτι του πολυδιάστατου μοντέλου ή κύβου, που παρουσιάζει τις πωλήσεις ανά εβδομάδα που παρακολούθηθηκαν. Όπως φαίνεται

στον Πίνακα 2.4 μπορούμε να αυξήσουμε το επίπεδο της ιεραρχίας ομαδοποιώντας είτε το Χρόνο σε κλίμακες, δηλαδή από εβδομάδες σε μήνες, είτε τις Πωλήσεις, δηλαδή δημιουργία τριών κλιμάκων Υψηλή-Μέση-Χαμηλή.

Πωλήσεις	2.522	5.782	10.634	26.652
Εβδομάδα1	1	0	3	0
Εβδομάδα2	0	2	0	1
Εβδομάδα3	1	0	3	0
Εβδομάδα4	0	2	0	1

Πίνακα 2.4

Στην πρώτη περίπτωση, δηλαδή εάν αλλάζαμε την κλίμακα του χρόνου το αποτέλεσμα της συναθροιστικής ανόδου στο παράδειγμα του πίνακά μας θα ήταν όπως παρουσιάζεται στον Πίνακα 2.5.

Πωλήσεις	2.522	5.782	10.634	26.652
Μήνας1	2	4	6	2

Πίνακα 2.5

Στην δεύτερη περίπτωση, δηλαδή εάν αλλάζαμε την κλίμακα των Πωλήσεων ως Υψηλή (>15.000), Μέση(>6.000), Χαμηλή(<6.000), το αποτέλεσμα της συναθροιστικής ανόδου στο παράδειγμα του πίνακά μας θα ήταν όπως παρουσιάζεται στον Πίνακα 2.6.

Πωλήσεις	Υψηλή	Μέση	Χαμηλή
Εβδομάδα1	0	3	1
Εβδομάδα2	1	0	2
Εβδομάδα3	0	3	1
Εβδομάδα4	1	0	2

Πίνακα 2.6

Ακόμη, με τη συναθροιστική άνοδο θα μπορούσαμε να αφαιρέσουμε μια διάσταση, δηλαδή ανάλογα με τις κατηγορίες και τις υποκατηγορίες των διαστάσεων που υπάρχουν να αφαιρέσουμε ολοκληρωτικά κάποια κατηγορία ή υποκατηγορία. Κατά αυτόν τον τρόπο επιτυγχάνεται μείωση της διάστασης, δηλαδή να μη συμπεριλάβουμε στην απεικόνισή μας μια διάσταση, όπως την υποκατηγορία προϊόντων ή τα υποκαταστήματα μιας περιοχής.

• Αναλυτική Κάθοδος (Drill-Down)

Με τη λειτουργία της αναλυτικής καθόδου πραγματοποιείται το αντίστροφο της συναθροιστικής μεθόδου. Αυτό επιτυγχάνεται είτε με τη μείωση του επιπέδου μιας ιεραρχίας για μια διάσταση είτε αυξάνοντας το πλήθος των διαστάσεων. Για παράδειγμα εάν εφαρμοζόταν αναλυτική κάθοδος στον Πίνακα 2.4 τα αποτελέσματά της θα ήταν όπως παρουσιάζονται παρακάτω στον Πίνακα 2.7.

Πωλήσεις	2.522	5.782	10.634	26.652
-----------------	--------------	--------------	---------------	---------------

Πωλήσεις	2.522	5.782	10.634	26.652
Ημέρα1	0	0	1	0
Ημέρα2	0	0	0	0
Ημέρα3	0	0	1	0
Ημέρα4	0	0	0	0
Ημέρα5	1	0	0	0
Ημέρα6	0	0	1	0
Ημέρα7	0	0	0	0

Πίνακας 2.7

- **Τεμαχισμός (Slicing)**

Με τη λειτουργία του τεμαχισμού μας δίνεται η δυνατότητα να επιλέξουμε ένα πεδίο, δηλαδή μια διάσταση κατά την οποία θα διαμορφωθεί η απεικόνισή μας. Στο παράδειγμά μας, εάν επιλέξουμε για τον Πίνακα 2.7 ότι θέλουμε τις υψηλές πωλήσεις (>15.000) θα πάρουμε τα αποτελέσματα που παρουσιάζονται στον Πίνακα 2.8.

Πωλήσεις	Υψηλή
Ημέρα1	0
Ημέρα2	0
Ημέρα3	0
Ημέρα4	0
Ημέρα5	0
Ημέρα6	0
Ημέρα7	0
Ημέρα8	1

Πίνακας 2.8

- **Κάθετος Τεμαχισμός (Dicing)**

Με τη λειτουργία του κάθετου τεμαχισμού μας δίνεται η δυνατότητα επιλογής δύο ή περισσότερων διαστάσεων. Ακόμη, μπορούμε να φιλτράρουμε και την κλίμακα της διάστασης εφόσον υπάρχει. Για παράδειγμα, τις πωλήσεις μιας συγκεκριμένης μέρας (βλ. Πίνακα 2.9) ή το ύψος των πωλήσεων συγκεκριμένων ημερών (βλ. Πίνακα 2.10).

Πωλήσεις	2.522	5.782	10.634	26.652
Ημέρα4	0	0	0	0

Πίνακας 2.9

Πωλήσεις	Υψηλή	Μέση	Χαμηλή
Ημέρα3	0	1	0

Πωλήσεις	Υψηλή	Μέση	Χαμηλή
Ημέρα4	0	0	0
Ημέρα5	0	0	1

Πίνακας 2.10

Αφού είδαμε κάποιες από τις τεχνικές OLAP μπορούμε να κατανοήσουμε τις δυνατότητες που διαθέτουν σε έναν αναλυτή και το επίπεδο της ανάλυσης που μπορεί να επιτευχθεί. Για παράδειγμα σκεφτείτε να θέλαμε να αναλύσουμε την αποτελεσματικότητα μιας εκστρατείας μάρκετινγκ για μερικά προϊόντα ή για ορισμένες περιοχές. Θα ήταν εφικτό μετρώντας την αύξηση των πωλήσεων σε μια συγκεκριμένη περίοδο σύμφωνα με τη χρήση των τεχνικών. Ο συνδυασμός πολλών διαστάσεων μπορεί να μας δημιουργήσει πολλά ερωτήματα για τα δεδομένα μας, αλλά μας δίνει και σημαντική λεπτομέρεια στα αποτελέσματα της ανάλυσης μας και ιδιαίτερα στις αναφορές μας. Η διαχείριση και η δυνατότητα της επιλογής των απεικονίσεων μεταξύ πολλαπλών διαστάσεων οδηγεί την ανάλυση σε υψηλότερα επίπεδα, καθώς την καθιστά πιο στοχευόμενη και πολυπαραγοντική.

Η αποδοτικότητα και η αποτελεσματικότητα των τεχνικών όμως επηρεάζεται από τη δομή των δεδομένων, δηλαδή το μοντέλο δεδομένων που υποστηρίζουν. Βέβαια, πολλές από τις τεχνικές αυτές σε συνδυασμό με τη στατιστική μπορούν να υλοποιηθούν και σε σχεσιακές βάσεις δεδομένων, οι οποίες μπορεί να είναι δομημένες σε σχήμα αστερά ή σε σχήμα χιονονιφάδας. Μερικές από αυτές τις τεχνικές θα τις δούμε στην πράξη και στο πειραματικό μέρος αυτής της εργασίας.

2.3.5 Προϋποθέσεις μιας Αποθήκης Δεδομένων

Τα επιχειρηματικά δεδομένα αποτελούν περιουσιακά στοιχεία σε έναν οργανισμό, αλλά η σωστή δόμησή και διαχείρισή τους αποτελεί θεμέλιο για τη μετέπειτα μετατροπή τους σε κέρδος σε έναν οργανισμό. Ο σχεδιασμός μιας αποθήκης δεδομένων θα πρέπει να λαμβάνει υπόψη κάποιες παραμέτρους, ώστε να καλύπτει σημαντικές προϋποθέσεις για την ομαλή λειτουργία των διαδικασιών που διαθέτει. Μερικές από αυτές είναι οι εξής:

- Η πρόσβαση στα δεδομένα μιας αποθήκης δεδομένων δε θα πρέπει να αποτελεί μια πολύπλοκη διαδικασία. Το περιεχόμενο των δεδομένων θα πρέπει να μπορεί να κατανοηθεί από τους χρήστες, ώστε να έχουν τη δυνατότητα να το παραμετροποιήσουν και να το χρησιμοποιήσουν σε πολλαπλούς συνδυασμούς, όπως στη διαδικασία ενός τεμαχισμού (slicing) ή ενός φίλτρου επιλογής συγκεκριμένων διαστάσεων (dicing). Ακόμη, οι τεχνικές που θα χρησιμοποιήσουν οι χρήστες θα πρέπει να είναι απλές, ώστε να επιτρέπουν την άμεση έκβαση των αποτελεσμάτων.
- Οι πληροφορίες που παρουσιάζει μια αποθήκη δεδομένων θα πρέπει να είναι αξιόπιστες. Η συγκέντρωση των δεδομένων από τις διαφορετικές πηγές και η διανομή τους θα πρέπει να πραγματοποιείται με προσοχή, που σημαίνει ότι θα ομαδοποιούνται ορθά σύμφωνα με την επιχειρηματική διαδικασία που αφορούν. Θα διατηρούν μια σχέση μεταξύ τους και θα διακρίνονται από τη μοναδικότητά τους, δηλαδή δε θα υπάρχουν παρόμοια δεδομένα για τον ίδιο σκοπό. Επίσης, θα πρέπει να αποτελούν υψηλής ποιότητας δεδομένα, δηλαδή δεδομένα που μπορούν να υπολογιστούν και δεν είναι ελλιπή. Αυτή η προϋπόθεση αποτελεί βασικό παράγοντα για τη συνέπεια μιας αποθήκης δεδομένων απέναντι στους χρήστες.

- Οι αλλαγές των επιχειρηματικών συνθηκών δε θα πρέπει να επηρεάσουν το περιεχόμενο και τις διαδικασίες μιας αποθήκης δεδομένων. Κατά το σχεδιασμό της θα πρέπει να ληφθεί υπόψη οποιαδήποτε μελλοντική πιθανή αλλαγή και να καθοριστεί η διαδικασία προσαρμογής της σε αυτές τις αλλαγές. Επιπλέον, η διαδικασία προσαρμογής της μπορεί να εγκυμονεί κινδύνους για το ήδη υπάρχον ιστορικό των δεδομένων της, οπότε χρειάζεται προσοχή. Για παράδειγμα, σκεφτείτε μια αλλαγή σε κατηγορηματικές μεταβλητές, όπου θα πρέπει να τροποποιηθούν σωστά, ώστε να καταγραφούν κατάλληλα.
- Η ασφάλεια και η προστασία των δεδομένων θα πρέπει να καλύπτεται από μια αποθήκη δεδομένων. Εφόσον περιλαμβάνει σημαντικές πληροφορίες για τους πόρους ενός οργανισμού, θα πρέπει να διαθέτει τις κατάλληλες ρυθμίσεις που αφορούν τη δυνατότητα πρόσβασης στα περιεχόμενά της. Για παράδειγμα, δεδομένα που αφορούν τη διαδικασία των πωλήσεων ή την κοστολόγηση κάποιων προϊόντων δε θα είναι προσβάσιμα σε κάθε χρήστη. Ο έλεγχος δηλαδή της πρόσβασης σε εμπιστευτικά δεδομένα αποτελεί κύριο ζήτημα κατά το σχεδιασμό μιας αποθήκης δεδομένων.
- Η υποστήριξη και η βελτίωση μιας διαδικασίας λήψης αποφάσεων θα πρέπει να είναι εφικτή. Αυτό σημαίνει πως η αποθήκη πρέπει να διαθέτει δεδομένα που είναι ιδανικά ώστε να υποστηρίζουν τη λήψη μιας απόφασης. Έτσι, τα αποτελέσματα που θα παρουσιάσει, θα περιλαμβάνουν στοιχεία που έχουν αξία για τον οργανισμό και αποτελούν καλή “εικόνα” για τα περιεχόμενά της.

2.4 ΕΠΙΧΕΙΡΗΜΑΤΙΚΗ ΕΥΦΥΙΑ ΚΑΙ ΕΞΟΡΥΞΗ ΔΕΔΟΜΕΝΩΝ

Όπως αναφέρθηκαν και σε προηγούμενη ενότητα, τα κύρια μέρη που αποτελείται ένα Σύστημα Επιχειρηματικής Ευφυΐας είναι πέντε. Για την καλύτερη κατανόηση των σταδίων που ακολουθεί ένα τέτοιο σύστημα θα αναλύσουμε παραπάνω τις διαδικασίες-κατηγορίες των εργαλείων που χρησιμοποιεί με βάση την Εξόρυξη Δεδομένων.

Η Εξόρυξη Δεδομένων (data mining) ή Εξόρυξη γνώσης αποτελεί μια επαναληπτική διαδικασία κατά την οποία η εξέλιξή της ορίζεται από τα αποτελέσματα της ανακάλυψής της, είτε με αυτόματες είτε με χειροκίνητες μεθόδους. Είναι πολύ χρήσιμη στη διερευνητική ανάλυση, κατά την οποία δεν υπάρχουν προκαθορισμένες αντιλήψεις για το τελικό αποτέλεσμα. Οι τεχνικές της αποσκοπούν στην γρήγορη αναζήτηση νέων, πολύτιμων και μη κερδοσκοπικών πληροφοριών σε μεγάλους όγκους δεδομένων με σκοπό την ανακάλυψη κρυμμένων χαρακτηριστικών. Οπότε μπορούμε να πούμε ότι η εξόρυξη γνώσης είναι μια διαδικασία που με τη συνεργασία της ανθρώπινης γνώσης και της τεχνολογίας μπορεί να εξάγει κρυφές πληροφορίες από βάσεις δεδομένων μεγάλου όγκου.

Σύμφωνα με τους Fayyad & Frawley, η Εξόρυξη Δεδομένων ορίζεται ως η διαδικασία ανεύρεσης ενδιαφέρουσας γνώσης, όπως τα πρότυπα, οι συσχετίσεις, οι αλλαγές, οι ανωμαλίες και οι σημαντικές δομές, από μεγάλες ποσότητες δεδομένων που αποθηκεύονται σε βάσεις δεδομένων, αποθήκες δεδομένων ή άλλα αποθετήρια πληροφοριών. Η διαθεσιμότητα αυτού του μεγάλου όγκου πληροφοριών σε ηλεκτρονική μορφή και η ανάγκη εκμετάλλευσής της χρησιμότητάς τους σε εφαρμογές ή σε συστήματα, όπως συστήματα υποστήριξης αποφάσεων ή συστήματα

διαχείρισης εφοδιαστικής αλυσίδας, έχουν αυξήσει σημαντικά το ενδιαφέρον και την προσοχή στην εξόρυξη δεδομένων.

Στην πράξη, οι δυο πρωταρχικοί στόχοι της εξόρυξης δεδομένων τείνουν στην πρόβλεψη και την περιγραφή. Η πρόβλεψη περιλαμβάνει την εκμετάλλευση κάποιων ορισμένων μεταβλητών ή πεδία με σύνολα στοιχείων, ενώ η περιγραφή επικεντρώνεται στην εξερεύνηση προτύπων που πηγάζουν από τα δεδομένα και είναι εφικτό να ερμηνευτούν από τον άνθρωπο.

Επομένως, οι δραστηριότητές της μπορούν να κατηγοριοποιηθούν σε:

-Εξόρυξη Δεδομένων Πρόβλεψης, όπου δημιουργεί το μοντέλο του συστήματος που περιγράφεται από το σύνολο των δεδομένων.

-Περιγραφική Εξόρυξη Δεδομένων, όπου παράγει νέες πληροφορίες με βάση την τροφοδότησή της από το διαθέσιμο σύνολο των δεδομένων.

Η γενική πειραματική διαδικασία της Εξόρυξης Γνώσης όμως μπορεί να χωριστεί στα παρακάτω βήματα:

-Δήλωση του προβλήματος και διατύπωση της υπόθεσης

Η πλειοψηφία των μελετών μοντελοποίησης των δεδομένων πραγματοποιούνται σε καθορισμένο τομέα εφαρμογής. Βέβαια, εξειδικευμένες γνώσεις και εμπειρία αποτελούν απαραίτητα εφόδια για τη δήλωση ενός προβλήματος. Η επένδυση στην τεχνική της εξόρυξης δεδομένων από μόνη της, δίχως τη σαφή-ορθή δήλωση του προβλήματος συνήθως δεν φέρει τα επιθυμητά αποτελέσματα. Σε αυτό το βήμα θα πρέπει ο υπεύθυνος-μοντελιστής να καθορίσει ένα σύνολο μεταβλητών κι αν είναι εφικτό την γενική μορφή της εξάρτησης ως αρχική υπόθεση. Απαιτείται δηλαδή στενή σχέση της γνώσης του μοντελιστή και της γνώσης του υπεύθυνου εφαρμογής, πράγμα που δεν σταματά στο πέρας του βήματος, αλλά διατηρείται και στα επόμενα βήματα.

-Συλλογή δεδομένων

Σε αυτό το βήμα περιλαμβάνεται ο τρόπος κατά τον οποίο συλλέγονται και παράγονται τα δεδομένα. Υπάρχουν δύο περιπτώσεις κατά τη συλλογή. Η μία μπορεί να πραγματοποιείται υπό τον έλεγχο του μοντελιστή (πείραμα σχεδιασμού), ενώ στην άλλη ο εμπειρογνώμονας δε μπορεί να έχει πρόσβαση στη διαδικασία, γνωστή ως παρατηρητική προσέγγιση. Η κατανομή δειγματοληψίας όμως αποτελεί βασικό παράγοντα για τα επόμενα βήματα, διότι συμβάλλει στη μοντελοποίηση και κατά συνέπεια στην ερμηνεία των αποτελεσμάτων, του μοντέλου και των δεδομένων που επιλέχθηκαν για χρήση.

-Προ-επεξεργασία δεδομένων

Κατά την παρατηρητική προσέγγιση, που αφορά τον τρόπο συλλογής δεδομένων που πραγματοποιείται από υπάρχουσες βάσεις δεδομένων, αποθήκες δεδομένων, η προ-επεξεργασία δεδομένων είναι απαραίτητη και συνήθως περιλαμβάνει κυρίως δύο διαδικασίες:

Ανίχνευση αποκλίσεων και αφαίρεση

Όταν υπάρχουν δεδομένα που δε συμφωνούν με τις περισσότερες παρατηρήσεις. Για παράδειγμα, μία λανθασμένη αριθμητική τιμή σε ένα σύνολο δεδομένων λόγω λανθασμένης μέτρησης. Αποτελεί σφάλμα και εάν το πλήθος τέτοιων τιμών είναι σημαντικό, εξίσου σημαντική θα είναι και η επιρροή στο μοντέλο αργότερα.

Κλιμάκωση, κωδικοποίηση και επιλογή χαρακτηριστικών

Αυτό συμβαίνει όταν κάποια δεδομένα έχουν διαφορετικά χαρακτηριστικά γνωρίσματα μεταξύ τους, δηλαδή δεν έχουν τηματοποιηθεί σωστά ή κατά το βέλτιστο τρόπο. Για παράδειγμα, ένα χαρακτηριστικό με εύρος $[0,1]$ κι ένα άλλο με $[-100,100]$ δε θα έχουν το ίδιο βάρος κατά την εφαρμογή της τεχνικής, αλλά θα ασκήσουν ανάλογη επιρροή στα αποτελέσματα της εξόρυξης δεδομένων. Με την κωδικοποίηση επιτυγχάνεται ένας μικρότερος αριθμός πληροφοριακών χαρακτηριστικών για τη μετέπειτα μοντελοποίηση.

-Εκτίμηση μοντέλου

Σε αυτή τη φάση πραγματοποιείται η επιλογή και η εφαρμογή της κατάλληλης τεχνικής εξόρυξης δεδομένων. Αυτό το βήμα είναι πολύπλοκο και στην πράξη βασίζεται στη βέλτιστη επιλογή του μοντέλου μεταξύ πολλών.

-Ερμηνεία του μοντέλου και εξαγωγή συμπερασμάτων

Στο τελικό αυτό βήμα, τα μοντέλα εξόρυξης πρέπει να είναι εύκολο να ερμηνευτούν ώστε να βοηθήσουν στη λήψη αποφάσεων. Εάν δεν είναι ερμηνεύσιμα δε μπορούν οι υπεύθυνοι ενός οργανισμού να στηρίζουν τις αποφάσεις τους σε πολύπλοκα μοντέλα. Η ακρίβεια όμως μειώνεται καθώς η απλότητα των μοντέλων αυξάνεται, διότι για να επιτύχουμε εξαιρετικά ακριβή συμπεράσματα χρησιμοποιούμε μοντέλα μεγάλης διαστάσεως. Από την πλευρά του χρήστη όμως είναι δύσκολο να ερμηνευθούν εκατοντάδες αριθμητικά αποτελέσματα, γι' αυτό η διαδικασία αυτή θα πρέπει να αποτελεί ξεχωριστό έργο.

2.4.1 Τεχνικές Εξόρυξης Δεδομένων

Για να επιτύχει τους στόχους της η Εξόρυξη Δεδομένων, δηλαδή να αντλήσει κρυμμένες χρήσιμες πληροφορίες, να εντοπίσει κάποιες τάσεις που αφορούν τη συμπεριφορά των δεδομένων, να ερμηνεύσει διάφορα στοιχεία κι ευρύτερα να μελετήσει ώστε να εξάγει ερμηνεύσιμα αποτελέσματα χρησιμοποιεί κάποιες μεθόδους-τεχνικές. Μερικές από αυτές είναι οι παρακάτω:

• Κατηγοριοποίηση ή Ταξινόμηση (Classification)

Κατά τη διαδικασία της κατηγοριοποίησης δημιουργείται ένα μοντέλο ώστε να προβλέψει τις μελλοντικές συμπεριφορές των πελατών με τη βοήθεια της ταξινόμησης των αρχείων στις βάσεις δεδομένων ανά κατηγορίες λαμβάνοντας υπόψη κάποια καθορισμένα κριτήρια [18]. Πραγματοποιείται δηλαδή οργάνωση ενός πλήθους στοιχείων σε ένα σύνολο (objects), το οποίο μπορεί να ερμηνευθεί από ένα σύνολο χαρακτηριστικών (attributes), σε κλάσεις (classes), οι οποίες είναι καθορισμένες από πριν. Σε αυτήν τη διαδικασία χρησιμοποιούνται μέθοδοι μάθησης με επίβλεψη. Οι τεχνικές ταξινόμησης χρησιμοποιούν δεδομένα, που είναι ήδη κατηγοριοποιημένα σε κλάσεις, με τη μορφή ενός συνόλου εκπαίδευσης (training set). Ο αλγόριθμος εξετάζει αυτές τις προκαθορισμένες κατηγορίες και στη συνέχεια αφού αναγνωρίσει κάποιο γνώρισμα των αντικειμένων υλοποιεί νέα ταξινόμηση των δεδομένων στις αρμόδιες κλάσεις. Επομένως, η τεχνική αυτή διαχειρίζεται την τοποθέτηση ενός στοιχείου με βάση τις υπάρχουσες κατηγορίες. Ο βασικός σκοπός είναι να δημιουργηθεί ένα μοντέλο όπου θα έχει τη δυνατότητα να ταξινομήσει δεδομένα, τα οποία δεν διατηρούν κάποια κατηγορία. Για να γίνει εφικτό αυτό, χρησιμοποιούνται διάφορες τεχνικές. Μερικές από αυτές είναι τα δέντρα απόφασης, τα νευρωνικά δίκτυα και οι αλγόριθμοι κανόνων.

• Συσταδοποίηση ή Ομαδοποίηση (Clustering)

Οι μέθοδοι ομαδοποίησης χωρίζουν τα αντικείμενα ή κάποιο πλήθος αντικειμένων, σύμφωνα με καθορισμένα κριτήρια, σε ομάδες με σκοπό τα αντικείμενα που ανήκουν σε μια ομάδα να είναι πιο παρόμοια μεταξύ τους ή να έχουν παρόμοια συμπεριφορά σε σύγκριση με άλλα αντικείμενα που ανήκουν σε άλλη ομάδα [19]. Μια σημαντική διαφορά της συσταδοποίησης από την κατηγοριοποίηση είναι ότι δεν υπάρχει γνώση που αφορά τα δεδομένα, δηλαδή τον τρόπο που δομούνται και τις κατηγορίες που μπορεί να έχουν. Οι αλγόριθμοι θα καθορίσουν την ομοιότητα μέσα σε μια ομάδα, προσπαθώντας να τη μεγιστοποιήσουν ανάμεσα στα αντικείμενα (intra-class similarity) και να την ελαχιστοποιήσουν μεταξύ των διαφορετικών ομάδων που απαρτίζουν άλλα αντικείμενα (inter-class similarity). Σύμφωνα με το Larose (2005), το σύμπλεγμα (cluster) δεν εκτελεί ταξινόμηση, εκτίμηση ή πρόβλεψη για την αξία των αντικειμένων, αλλά πραγματοποιεί την κατανομή ομογενοποιημένων υπό-ομάδων. Μερικές από τις βασικές μεθόδους συσταδοποίησης είναι:

▪ Διαχωριστικές μέθοδοι (partitioning methods)

Οι μέθοδοι ομαδοποίησης με βάση το διαχωρισμό είναι ευέλικτες μέθοδοι που στηρίζονται στην επαναλαμβανόμενη εγκατάσταση των σημείων των αντικειμένων μεταξύ των συμπλεγμάτων (clusters). Σε κάθε επανάληψη, οι επαναληπτικοί αλγόριθμοι μειώνουν την αξία της συνάρτησης του κριτηρίου έως τη σύγκλιση. Κάθε αντικείμενο ανήκει σε μία συστάδα, κάτι που μπορεί να αλλάξει σε μία επόμενη επανάληψη ή στην αλλαγή του κριτηρίου. Ο αριθμός των συστάδων μπορεί να καθοριστεί απ' το χρήστη και αλλάζοντας το κριτήριο της ομαδοποίησης είναι δυνατό να δημιουργηθούν ισχυρές μέθοδοι ομαδοποίησης που είναι λιγότερο ευαίσθητες σε ελλιπή και μη-κατάλληλα δεδομένα. Ο πιο γνωστός αλγόριθμος διαχωριστικής ομαδοποίησης είναι ο k-Means.

▪ Ιεραρχικές μέθοδοι (hierarchical methods)

Οι ιεραρχικές μέθοδοι είναι είτε συσσωρευτικές (agglomerative) είτε διαιρετικές (divisive). Έστω ότι θέλουν να ομαδοποιηθούν n αντικείμενα, οι συσσωρευτικές μέθοδοι ξεκινούν με n συστάδες, δηλαδή κάθε αντικείμενο αποτελεί μια συστάδα ξεχωριστά και σε κάθε βήμα επιλέγονται και συγχωνεύονται δύο ομάδες με γνώμονα την ομοιότητά τους. Αυτή η διαδικασία συνεχίζεται έως ότου όλα τα αντικείμενα συγκεντρωθούν σε μία ομάδα (bottom up προσέγγιση). Από την άλλη πλευρά, οι διαιρετικές μέθοδοι ξεκινούν με μία ενιαία συστάδα που περιλαμβάνει όλα τα αντικείμενα. Σε κάθε βήμα, επιλέγεται ένα σύμπλεγμα με βάση τη διαφορετικότητα μεταξύ των δεδομένων και χωρίζονται σε δύο υπό-ομάδες. Αυτή η διαδικασία επαναλαμβάνεται έως ότου όλα τα αντικείμενα να αποτελούν μια συστάδα, δηλαδή βάσει του παραδείγματος n συστάδες (top down προσέγγιση).

▪ Μέθοδοι βασισμένες σε μοντέλα (model-based methods)

Οι μέθοδοι συσταδοποίησης που είναι βασισμένες σε μοντέλα, χρησιμοποιούν μοντέλα ώστε να εντοπίσουν τα διάφορα αντικείμενα που υπάρχουν σε κάθε σύμπλεγμα (cluster). Θεωρούν πως κάθε σύμπλεγμα εκφράζεται με ένα μαθηματικό μοντέλο. Σύμφωνα με τους Yeung και Fraley, η ομαδοποίηση βάσει μοντέλων υποθέτει ότι τα δεδομένα δημιουργήθηκαν από ένα πεπερασμένο μίγμα κατανομών πιθανοτήτων, όπως πολλαπλές μεταβλητές κανονικής κατανομής [30]. Αυτές οι μέθοδοι έχουν ως στόχο τη βελτίωση των σχέσεων που συνδέουν τα δεδομένα και τα μοντέλα. Μία τέτοια μέθοδο αποτελούν οι Αυτό-οργανωμένοι Χάρτες (Self-Organizing Maps), οι οποίοι ανήκουν στα νευρωνικά δίκτυα.

• Κανόνες Συσχέτισης (Association Rule Mining)

Σύμφωνα με τον Larose (2005), η διαδικασία συσχέτισης εντοπίζει ποια χαρακτηριστικά "πηγαίνουν μαζί". Δηλαδή, ο σκοπός αυτής της διαδικασίας είναι να ανακαλύψει ποια στοιχεία μέσα σε ένα αρχείο διατηρούν κάποια σχέση ή σύνδεση. Κατά κάποιο τρόπο συνοψίζει τα δεδομένα. Ο Baker (2010), υποστηρίζει πως με τους κανόνες συσχέτισης πραγματοποιείται η <<ανακάλυψη του κανόνα if-then>> που σημαίνει ότι αν βρεθεί η τιμή μιας μεταβλητής, η τιμή μιας άλλης μεταβλητής θα έχει συγκεκριμένη αξία [22]. Για παράδειγμα, οι πελάτες που προτιμούν ένα συγκεκριμένο προϊόν μπορεί επίσης να επιλέξουν κάποιο άλλο που να έχει σχέση-σύνδεση με αυτό που προτιμούν. Οι αλγόριθμοι Apriori, CARMA και GRI χρησιμοποιούνται στους κανόνες συσχέτισης.

2.4.2 Δέντρα Απόφασης

Τα δέντρα αποφάσεων αποτελούν ισχυροί αλγόριθμοι ταξινόμησης. Σύμφωνα με τον Yin (2011), ορίζει το δέντρο απόφασης ως μια "συστηματική μέθοδο που χρησιμοποιεί ένα διάγραμμα δέντρων, το οποίο περιλαμβάνει κόμβους και κλάδους, οι κόμβοι συνδέονται με κλάδους, ο χρόνος κυλάει από αριστερά προς τα δεξιά και κάθε ράντσο αντιπροσωπεύει μια απόφαση ή ένα πιθανό γεγονός" [31]. Όπως υποδηλώνει το όνομα, η τεχνική αυτή χωρίζει αναδρομικά τις παρατηρήσεις σε κλάδους για να κατασκευάσει ένα δέντρο, το οποίο θα βοηθήσει στη βελτίωση της ακρίβειας στην πρόβλεψη. Κατά αυτόν τον τρόπο, χρησιμοποιούν μαθηματικούς αλγορίθμους, οι οποίοι αναγνωρίζουν μια μεταβλητή και το αντίστοιχο όριο για τη μεταβλητή, το οποίο χωρίζει την παρατήρηση εισόδου σε δύο ή περισσότερες υπο-ομάδες. Αυτή η διαδικασία πραγματοποιείται επαναληπτικά έως ότου κατασκευασθεί το πλήρες δέντρο. Για την καλύτερη κατανόηση ας δούμε παραδειγματικά τη διαδρομή που ακολουθεί ένα νέο αντικείμενο ώστε να ταξινομηθεί. Ξεκινώντας από τη ρίζα, δηλαδή τον αρχικό κόμβο, πραγματοποιείται εξέταση των χαρακτηριστικών του και προσδιορίζεται το νέο αντικείμενο, ανάλογα με τα αποτελέσματα της εξέτασης, σε ποιους εσωτερικούς κόμβους θα τοποθετηθεί. Η ίδια διαδικασία επαναλαμβάνεται μέχρι το νέο αντικείμενο να τοποθετηθεί σε ένα από τα φύλλα του δέντρου, το οποίο αποτελεί τελικό κόμβο κι αντιστοιχεί στην κλάση που θα ταξινομηθεί το νέο αυτό αντικείμενο. Έτσι, μπορούμε να πούμε πως για την κατηγοριοποίηση με τη χρήση δέντρων αποφάσεων πρέπει αρχικά να πραγματοποιηθούν δύο λειτουργίες, οι οποίες αποτελούν θεμέλιο για τη δημιουργία ενός δέντρου απόφασης. Η πρώτη αφορά την αξιολόγηση των χαρακτηριστικών κάθε παρατήρησης, ώστε να χωριστεί όσο το δυνατόν καλύτερα σε υπό-κατηγορίες. Η δεύτερη αφορά την υλοποίηση αυτού του διαχωρισμού που εξετάζεται στην πρώτη λειτουργία, δηλαδή τη δημιουργία υπό-ομάδων. Ακόμη, τα δέντρα απόφασης περιλαμβάνουν διάφορους αλγόριθμους για την ταξινόμηση, μερικούς από τους οποίους θα δούμε παρακάτω.

Επιπλέον, τα δέντρα αποφάσεων έχουν αρκετά πλεονεκτήματα έναντι των διαδικασιών ταξινόμησης που χρησιμοποιούν για παράδειγμα τη μέγιστη πιθανότητα. Συγκεκριμένα, τα δέντρα αποφάσεων δεν απαιτούν τη δημιουργία υποθέσεων που να αφορούν τις κατανομές των δεδομένων εισόδου και είναι μη παραμετρικά. Ακόμη ένας λόγος που ξεχωρίζουν στις μεθόδους κατηγοριοποίησης είναι ότι η ερμηνεία τους είναι εύκολη και οι κανόνες που ακολουθούν μπορούν να κατανοηθούν από κάθε χρήστη. Επιπλέον, μπορούν να χειριστούν αριθμητικές και κατηγορικές μεταβλητές κι ευρύτερα δεδομένα που έχουν πλήθος γνωρισμάτων. Ιδιαίτερα, έχουν

τη δυνατότητα να διαχειριστούν δεδομένα που δεν έχουν απαραίτητα γραμμική σχέση μεταξύ τους και η απόδοσή τους σε μεγάλο όγκο δεδομένων είναι αρκετά καλή. Πέρα από τον μεγάλο όγκο δεδομένων, η απουσία ή η μη καταλληλότητα των δεδομένων δεν αποτελούν εμπόδιο στη δραστηριότητα των δέντρων αποφάσεων, διότι υπάρχουν τεχνικές που προσαρμόζουν κάθε περίπτωση στους κανόνες που αυτά ακολουθούν, όπως η μετατροπή κάποιων αριθμητικών μεταβλητών σε κατηγορικές μεταβλητές.

Ωστόσο, οι αλγόριθμοι των δέντρων αποφάσεων έχουν και τα μειονεκτήματά τους, διότι δεν παρέχουν την έξοδο πλήθους χαρακτηριστικών και δεν είναι σταθερά. Κάποιες παραλλαγές των μεταβλητών μπορεί να επηρεάσουν την επιλογή του σημείου τοποθέτησής τους μέσα στο δέντρο. Έτσι, η επιρροή στην ευρύτερη δομή του δέντρου μπορεί να είναι τόσο σημαντική ώστε να οδηγήσει σε εσφαλμένη εξαγωγή των συμπερασμάτων, διότι η επιλογή των γνωρισμάτων για την ταξινόμηση αποτελεί καθοριστικό παράγοντα για τα αποτελέσματα ενός δέντρου απόφασης. Ακόμη, η πολυπλοκότητα των αριθμητικών μεταβλητών και το μέγεθος του όγκου τους μπορεί να επηρεάσει σημαντικά τη διαδικασία του διαχωρισμού τους σε σημείο που να καθιστά αδύνατη την τμηματοποίησή τους.

2.4.2.1 IDE Αλγόριθμος

Ο ID3 αλγόριθμος δημιουργήθηκε από τον Quinlan (1986) και η διαδικασία που ακολουθεί ώστε να δομηθεί ένα δέντρο απόφασης είναι αρχικά να ελέγχει εάν όλες οι περιπτώσεις ενός κόμβου είναι από την ίδια κλάση. Εάν είναι, τότε το δέντρο είναι απλά ένας κόμβος απαντήσεων, ο οποίος περιέχει το όνομα της κλάσης. Σε αντίθετη περίπτωση, ο αλγόριθμος μετρά τον αριθμό των θετικών και των αρνητικών χαρακτηριστικών για κάθε τιμή της κάθε περίπτωσης. Στη συνέχεια εκτιμά ένα σκορ για κάθε χαρακτηριστικό, επιλέγει το χαρακτηριστικό με το χαμηλότερο σκορ, όπου το χρησιμοποιεί δοκιμαστικά στον κόμβο, και δημιουργεί αναδρομικά ένα δέντρο για κάθε υποσύνολο των περιπτώσεων με την ίδια τιμή του επιλεγμένου χαρακτηριστικού δοκιμής [22]. Εάν ένας κόμβος διατηρεί μόνο ένα πιθανό χαρακτηριστικό, τότε δεν είναι απαραίτητο να υπολογίσει το σκορ. Δηλαδή για κάθε κόμβο του δέντρου που δεν τερματίζεται επιλέγει ακολουθιακά ένα χαρακτηριστικό και για κάθε τιμή του δημιουργεί έναν κλάδο. Η διαδικασία αυτή επαναλαμβάνεται έως ότου όλες οι περιπτώσεις ενός κόμβου ανήκουν στην ίδια κλάση ή όταν δεν εντοπίζεται κάποιο χαρακτηριστικό που να διαχωρίζει τα στοιχεία ή όταν δεν υπάρχουν υπό-ομάδες του συνόλου με παρατηρήσεις.

2.4.3 Νευρωνικά Δίκτυα

Η έμπνευση για τα νευρωνικά δίκτυα προέκυψε από την αναγνώριση της σύνδεσης μεταξύ των πολύπλοκων συστημάτων μάθησης στον εγκέφαλο των ζώων με τα σύνολα των νευρώνων. Μπορεί ένας καθορισμένος νευρώνας να είναι εύκολος στη δομή του, αλλά τα πυκνά δίκτυα των διασυνδεδεμένων νευρώνων θα μπορούσαν να επιτύχουν την υλοποίηση σύνθετων μαθησιακών καθηκόντων, όπως η ταξινόμηση και η αναγνώριση προτύπων. Τα τεχνητά νευρωνικά δίκτυα ή νευρωνικά δίκτυα έχουν ως βασικό σκοπό τη μίμηση του τύπου της μη γραμμικής μάθησης που εμφανίζεται στα δίκτυα των νευρώνων στη φύση [26]. Στην ουσία αποτελούν τεχνικές ανάλυσης όπου είναι ικανές να προβλέψουν καινούριες μεταβλητές από άλλες παρατηρήσεις ή από τις ίδιες τις μεταβλητές. Η βασική τους δομή αποτελείται από νευρώνες, οι οποίοι χρησιμοποιούν δενδρίτες για τη συλλογή των δεδομένων και συνδυάζοντας τις πληροφορίες εισόδου τους πραγματοποιούν τη μεταβίβαση αυτών

των δεδομένων σε άλλους συνδεδεμένους νευρώνες με τελικό στόχο μια απόκριση εξόδου, η οποία διοχετεύεται προς τα κάτω σε άλλους νευρώνες.

Ένα από τα πλεονεκτήματα της χρήσης των νευρωνικών δικτύων είναι ότι περιέχουν πολλούς κόμβους, οι οποίοι μπορούν να επεξεργάζονται πιο αποτελεσματικά τα δεδομένα μέσα στο δίκτυο και να πραγματοποιούν πιο άμεσα την ενημέρωση στο σύνολο των δεδομένων. Ακόμη, είναι σημαντικά ανθεκτικά σε θορυβώδη δεδομένα, αλλά σε σύγκριση με τα δένδρα αποφάσεων, που παράγουν κατανοητούς κανόνες, είναι σχετικά πιο δύσκολα στην ερμηνεία τους.

2.4.4 Αλγόριθμος *K-means*

Μια διαχωριστική μέθοδος συσταδοποίησης αποτελεί ο αλγόριθμος *K-means*, οποίος δημιουργήθηκε από τον MacQueen (1967) και αποτελεί αλγόριθμος μάθησης χωρίς επίβλεψη όπου επιλύει προβληματικές περιπτώσεις συγκέντρωσης χρησιμοποιώντας μια ποσοτική μέθοδο [28]. Όπως αναφέραμε παραπάνω με τη συσταδοποίηση σκοπός μας είναι να ομαδοποιήσουμε ένα σύνολο δεδομένων σε μικρότερα υποσύνολα. Οι δύο μέθοδοι με τους οποίους μπορούμε να το καταφέρουμε αυτό είναι ποσοτικά, χρησιμοποιώντας αριθμητικές μεταβλητές, και ποιοτικά, χρησιμοποιώντας κατηγορικές μεταβλητές. Η μέθοδος που ακολουθεί αυτός ο αλγόριθμος είναι αρχικά να ορίζει έναν προκαθορισμένο αριθμό συνόλων και να χρησιμοποιεί μαθηματικές πράξεις αλγοριθμικά για να ταξινομήσει τα δεδομένα στα σύνολα που ορίστηκαν εξ αρχής. Η εφαρμογή της μεθόδου μπορεί να χωριστεί σε κάποια βήματα, τα οποία είναι:

- Η αυθαίρετη επιλογή k αντικειμένων, τα οποία αποτελούν τον αριθμό των συνόλων, δηλαδή των συστάδων που θα προκύψουν. Αυτός ο αριθμός του πλήθους των συστάδων μπορεί να είναι και επιλογή του χρήστη.
- Ο υπολογισμός της διαφορετικότητας του αντικειμένου και του μέσου του συνόλου που αρχικά δημιουργήθηκε.
- Η κατανομή του αντικειμένου στο σύνολο (συστάδα), όπου η τιμή του μέσου της είναι η πλησιέστερη προς αυτό το αντικείμενο.
- Ο υπολογισμός του μέσου του συνόλου (συστάδα) σύμφωνα με τα αντικείμενα που τοποθετήθηκαν με σκοπό την ελαχιστοποίηση των αποστάσεων μεταξύ των συνόλων (συστάδων).

Πέρα από το αρχικό στάδιο, δηλαδή του προσδιορισμού του πλήθους των συστάδων, οι επόμενες διαδικασίες εκτελούνται επαναληπτικά στον αλγόριθμο έως ότου τα μέσα των συνόλων να μην μετακινούνται καθόλου. Στην ουσία ο αλγόριθμος αυτός εκτελείται έως ότου ελαχιστοποιηθεί κατά το βέλτιστο τρόπο το E της συνάρτησης (1) και συνήθως για τον υπολογισμό των αποστάσεων χρησιμοποιεί το τετράγωνο της Ευκλείδειας απόστασης:

$$E = \sum_{i=1}^k \sum_{x \in S_i} \|x - \mu_i\|^2 \quad (1)$$

k : ο αριθμός των συστάδων

x : το σημείο του αντικειμένου

μ_i : το σημείο του μέσου της συστάδας

Si: το σύνολο των συστάδων σύμφωνα με το k ($i=1,2,...,k$)

Ο αλγόριθμος K-means αποτελεί από τους λιγότερο πολύπλοκους αλγόριθμους και για αυτόν τον λόγο φέρει καλύτερα αποτελέσματα σε μεγάλα σύνολα δεδομένων. Από τη μία πλευρά είναι πιο λειτουργικός στην επεξεργασία μεγάλου πλήθους δεδομένων, αλλά από την άλλη πλευρά σε περιπτώσεις μη αριθμητικών μεταβλητών είναι δύσκολο να εκτελεστεί χάρη στο ότι δεν είναι εφικτός ο υπολογισμός του ελάχιστου των αποστάσεων σύμφωνα με την παραπάνω συνάρτηση (1). Ακόμη, η απλότητά του στα βήματα που εκτελεί τον καθιστά γρήγορο στη συσσώρευση ενός μεγάλου συνόλου των δεδομένων. Τέλος, ένα κρίσιμο σημείο του αποτελεί ο αρχικός καθορισμός των συστάδων, δηλαδή το k , όπου είναι βασικός παράγοντας της αποτελεσματικότητάς του και πολλές φορές υπολογίζεται με τη χρήση άλλων αλγορίθμων.

2.4.5 Αλγόριθμος Apriori

Η δημιουργία κανόνων συσχετίσεων συνήθως χωρίζεται σε δύο ξεχωριστά μέρη:

- Αρχικά, πραγματοποιείται η εύρεση όλων των αντικειμένων που εμφανίζονται συχνά σε μία βάση δεδομένων εφαρμόζοντας την μικρότερη υποστήριξη.
- Τα συχνά σύνολα που ανιχνεύθηκαν και ο ελάχιστος περιορισμός εμπιστοσύνης εξετάζονται για να σχεδιαστούν οι κανόνες.

Το πρώτο μέρος σε σύγκριση με το δεύτερο αποτελεί μια πιο πολύπλοκη διαδικασία, διότι όταν έχουμε μεγάλο σύνολο δεδομένων τα πιθανά ζεύγη συνόλων που προκύπτουν είναι σημαντικά πολλά. Παρά το γεγονός αυτό, είναι δυνατή η αποτελεσματική αναζήτηση κι ανίχνευση των στοιχειοσυνόλων χρησιμοποιώντας την ιδιότητα της υποστήριξης προς τα κάτω (αντιμονοτονικότητα), η οποία υποστηρίζει ότι για ένα συχνό σύνολο αντικειμένων ή ένα σπάνιο σύνολο αντικειμένων, όλα τα πιθανά υποσύνολά του πρέπει επίσης να είναι συχνά ή σπάνια αντίστοιχα. Ο αλγόριθμος Apriori χρησιμοποιεί αυτή την ιδιότητα και αποτελεί ο πρώτος που παρουσιάστηκε από τους Agrawal & Srikant (1994). Αρχικά, ο αλγόριθμος αυτός ανιχνεύει τα αντικείμενα που εμφανίζονται συχνά σύμφωνα με μια ελάχιστη υποστήριξη που καθορίζει ο χρήστης. Κατά τη διάρκεια αυτού του πρώτου βήματος κατασκευάζει τα υποψήφια 1-στοιχειοσύνολα. Στη συνέχεια, δημιουργεί τα συχνότερα 1-στοιχειοσύνολα κλαδεύοντας μερικά υποψήφια ζεύγη συνόλων, όπου οι τιμές υποστήριξης που έχουν είναι μικρότερες από αυτήν που έδωσε ο χρήστης αρχικά. Αφού εντοπίσει όλα τα συνηθισμένα 1-στοιχειοσύνολα, τα συνδέει σταδιακά μεταξύ τους κι αφαιρεί αυτά που εμφανίζονται σπάνια με τη διαδικασία του κλαδέματος. Έτσι, δημιουργεί τα συχνότερα 2-στοιχειοσύνολα και συνεχίζει επαναλαμβάνοντας τη διαδικασία έως ότου να μην είναι εφικτή η δημιουργία περισσότερων υποψήφιων στοιχείων [21].

Για να γίνει πιο κατανοητή η διαδικασία που ακολουθεί ο αλγόριθμος Apriori, ας δούμε ένα παράδειγμα. Έστω ότι έχουμε τα εξής στοιχεία αγορών σε μία βάση δεδομένων, όπου $\pi_1, \pi_2, \pi_3, \pi_4, \pi_5$ αποτελούν είδη προϊόντων (βλ. Πίνακα 2.11):

Αγορές
$\pi_1, \pi_2, \pi_3, \pi_4$
$\pi_1, \pi_2, \pi_3, \pi_4, \pi_5$

Αγορές
π_2, π_3, π_4
π_2, π_3, π_5
π_1, π_2, π_4
π_1, π_3, π_4
$\pi_2, \pi_3, \pi_4, \pi_5$
$\pi_1, \pi_3, \pi_4, \pi_5$
π_3, π_4, π_5
$\pi_1, \pi_2, \pi_3, \pi_5$

Πίνακας 2.11

Όπως είπαμε και παραπάνω ο αλγόριθμος σε πρώτο στάδιο θα βρει τη συχνότητα εμφάνισης των αντικειμένων, άρα θα μας δώσει το αποτέλεσμα του πίνακα 2.12

Αντικείμενα	Υποστήριξη
π_1	6
π_2	7
π_3	9
π_4	8
π_5	6

Πίνακας 2.12

Σε αυτό το σημείο θα θέσουμε το όριο της υποστήριξης, δηλαδή τη συχνότητα που εμφανίζεται ένα αντικείμενο. Θα δώσουμε ως ελάχιστο βαθμό το 4, ώστε να θεωρηθεί συχνή η εμφάνιση των αντικειμένων. Στη συνέχεια, θα δημιουργηθεί στοιχειοσύνολο με τα ζεύγη των αντικειμένων, δηλαδή δύο αντικειμένων (βλ. Πίνακα 2.13).

Αντικείμενα	Υποστήριξη
π_1, π_2	4
π_1, π_3	5
π_1, π_4	5
π_1, π_5	3
π_2, π_3	6
π_2, π_4	5
π_2, π_5	4
π_3, π_4	7
π_3, π_5	6

Αντικείμενα	Υποστήριξη
π_4, π_5	4

Πίνακας 2.13

Η τέταρτη γραμμή του πίνακα 2.13, δηλαδή αυτή με τα αντικείμενα π_1, π_5 και την υποστήριξη που ισούται με τρία θα αφαιρεθεί με τη διαδικασία του κλαδέματος, διότι δεν ικανοποιεί τη συνθήκη της υποστήριξης που θέσαμε παραπάνω.

Στη συνέχεια, θα δημιουργηθεί στοιχειοσύνολο με ζεύγη τριών αντικειμένων (βλ. Πίνακα 2.14), όπου και θα τερματίσει ο αλγόριθμος, διότι οποιοδήποτε άλλο σύνολο δεδομένων (αγορών), δηλαδή τεσσάρων αντικειμένων, δεν είναι τουλάχιστον τέσσερα η υποστήριξή τους στα δεδομένα μας.

Αντικείμενα	Υποστήριξη
π_1, π_3, π_4	4
π_2, π_3, π_4	4
π_2, π_3, π_5	4
π_3, π_4, π_5	4

Πίνακας 2.14

2.5 ΠΛΕΟΝΕΚΤΗΜΑΤΑ-ΜΕΙΟΝΕΚΤΗΜΑΤΑ

Τα Συστήματα Επιχειρηματικής Ευφυΐας παρέχουν πολλά οφέλη στους οργανισμούς που τα χρησιμοποιούν και επενδύουν σε αυτά, διότι αντλούν σημαντικές πληροφορίες από κάθε τμήμα τους ενισχύοντας την οργάνωσή τους και καθιστώντας τα μέλη που τους απαρτίζουν έτοιμα και σε εγρήγορση στις απρόσμενες αλλαγές των οικονομικών συνθηκών. Οι λύσεις όμως που προσφέρει η τεχνολογία εγκυμονεί και κινδύνους-περιορισμούς.

2.5.1 Πλεονεκτήματα Επιχειρηματικής Ευφυΐας

Μερικά από τα οφέλη που προσδίδουν τα Συστήματα Επιχειρηματικής Ευφυΐας παρατίθενται παρακάτω.

- Δυνατότητα συλλογής πληροφοριών από κάθε μέλος ενός οργανισμού και βέλτιστη αξιολόγηση αυτών με στόχο την εξαγωγή συμπερασμάτων που αφορούν είτε το εσωτερικό είτε το εξωτερικό περιβάλλον, όπως κατανόηση ή ομαδοποίηση πελατών, προμηθευτών και διάκριση ή διαχείριση ανταγωνιστών.
- Εντοπισμός κερδοφόρων πόρων στον οργανισμό, για παράδειγμα ένα ή και περισσότερα προϊόντα που ενδυναμώνουν τη σχέση τους με τους πελάτες τους και αξιολόγηση-εκτίμηση αυτών. Η αναγνώριση των αιτιών της πίστης αυτών των πελατών αποτελεί μεγάλο πλεονέκτημα για μια οργάνωση.
- Ανάλυση όλων των διαδικασιών που πραγματοποιούνται κι άμεση τροποποίηση αυτών κατά το βέλτιστο τρόπο. Για παράδειγμα, σε περίπτωση κάποιου προβλήματος ή βλάβης κατά την παραγωγική διαδικασία ενός προϊόντος, διατίθεται άμεση ενημέρωση και τα υπεύθυνα μέλη μπορούν να τροποποιήσουν τη διαδικασία ή ακόμη και τη στρατηγική που ακολουθούν.

- Βελτίωση στις επιχειρηματικές αποφάσεις, διότι η διοίκηση μπορεί να αντλεί τις κυριότερες πληροφορίες που χρειάζεται για να διαμορφώσει τους στόχους της [25].
- Επίτευξη ή διατήρηση ανταγωνιστικού πλεονεκτήματος. Η κατανόηση των πελατών και η εξυπηρέτησή τους αποτελεί θεμέλιο για τη διατήρηση της φήμης ενός προϊόντος. Ακόμη, η κατάλληλη διαχείριση των περιουσιακών στοιχείων μιας οργάνωσης μπορεί να οδηγήσει και σε αύξηση μεριδίου αγοράς.
- Αύξηση των εσόδων-κερδών[9]. Η καλύτερη πληροφόρηση μπορεί να βελτιώσει τη διαχείριση της εφοδιαστικής αλυσίδας κι αυτό καθιστά εφικτή τη μείωση του κόστους και στη συνέχεια την αύξηση των πωλήσεων.
- Αύξηση αποτελεσματικότητας και αποδοτικότητας ενός οργανισμού. Η άμεση κι η κατάλληλη ενημέρωση μεταξύ των τμημάτων σε έναν οργανισμό βελτιώνει την επικοινωνία μεταξύ των μελών που την απαρτίζουν κι αυτό έχει ως αποτέλεσμα την ορθότερη λειτουργία του.
- Εκμετάλλευση πιθανών ευκαιριών σε επιχειρηματικές αποφάσεις. Η κατανόηση της αγοράς μέσω της ανάλυσης καθιστά τα υπεύθυνα μέλη ικανά να διακρίνουν-εντοπίσουν κάποια πιθανή ευκαιρία και να είναι σε ετοιμότητα να την αξιοποιήσουν, είτε στο παρόν είτε στο μέλλον.
- Πρόβλεψη μελλοντικών αποτελεσμάτων με βάση τα τρέχοντα και τα ιστορικά δεδομένα χρησιμοποιώντας προγνωστική ανάλυση [10]. Επιτρέπει στους υπεύθυνους να μαθαίνουν από τα δεδομένα πώς να προβλέψουν μια μελλοντική συμπεριφορά των ατόμων.[11]. Επιπλέον, η κατανόηση των τάσεων της αγοράς αποτελεί πλεονέκτημα για μια οργάνωση, διότι της δίνεται η ευκαιρία να ανταποκριθεί σε οποιαδήποτε πρόκληση.
- Βελτίωση στην αξιοποίηση των Πληροφοριακών Συστημάτων που έχουν επενδύσει οι περισσότερες επιχειρήσεις. Επιπλέον, ο όγκος των πληροφοριών που διατηρείται και είναι άμεσα διαθέσιμος από κάθε σημείο μπορεί να αξιοποιηθεί με τα Συστήματα Επιχειρηματικής Ευφυΐας [12].

2.5.2 Μειονεκτήματα Επιχειρηματικής Ευφυΐας

Μερικοί από τους κινδύνους που εγκυμονούν τα Συστήματα Επιχειρηματικής Ευφυΐας παρατίθενται παρακάτω.

- Το κόστος υλικού για την υποστήριξη ενός Συστήματος Επιχειρηματικής Ευφυΐας αποτελεί τον πρώτο περιορισμό, διότι μπορεί να φτάσει σε σημαντικά μεγάλο βαθμό ιδιαίτερα εάν ένας οργανισμός δεν διατηρεί ήδη αρχείο δεδομένων [29].
- Το κόστος λογισμικού. Πολλοί οργανισμοί δεν μπορούν να ανταπεξέλθουν σε ένα τέτοιο άνοιγμα, για αυτό το λόγο περιορίζονται σε επιλογές συγκεκριμένων προγραμμάτων που κατά τη βούληση των υπεύθυνων μελών τους μπορούν να υποστηρίξουν και να τους εξελίξει.
- Ελλιπής ενημέρωση και γνώση από την πλευρά του προσωπικού. Ακόμη και αν υπάρχει ήδη κάποιο Πληροφοριακό Σύστημα εκ των προτέρων δεν εγγυάται ότι το υπεύθυνο προσωπικό μπορεί να διαχειριστεί οτιδήποτε νεότερο.
- Άρνηση από την πλευρά του προσωπικού, λόγω του προηγούμενου μειονεκτήματος ή λόγω του ότι ακολουθεί συγκεκριμένες οδηγίες σύμφωνα με τα τμήματα και τις λειτουργίες ενός οργανισμού, όπου θα επηρεαστούν στον ερχομό ενός νέου συστήματος.
- Κόστος εκπαίδευσης προσωπικού ή πρόσληψη νέου εξειδικευμένου σε Συστήματα Ευφυΐας. Αυτό περιλαμβάνει κι άλλο μειονέκτημα, διότι απαιτείται να δαπανηθεί πολύτιμος χρόνος μέχρι να επιτευχθεί η ορθή λειτουργία του συστήματος.

- Αδυναμία στην ενοποίηση τους με άλλα συστήματα. Η πλειοψηφία των οργανώσεων σήμερα χρησιμοποιούν διάφορα Πληροφοριακά Συστήματα όπου η έκδοσή τους ή η τροποποίηση που τους έχει πραγματοποιηθεί να καθιστά αδύνατη τη συμβατότητά τους με άλλα, όπως τα Συστήματος Επιχειρηματικής Ευφυΐας.
- Καταλληλότητα όγκου και ποιότητας δεδομένων. Όταν το σύστημα φορτωθεί με εσφαλμένα, ημιδομημένα ή δεδομένα που δεν έχουν καμία συνοχή μεταξύ τους, θα αδυνατεί να λειτουργήσει ομαλά και να επιφέρει αποτελέσματα λόγω της κακής πληροφόρησης που πραγματοποιήθηκε εξ αρχής.
- Οι επιτυχίες σε έργα τέτοιων συστημάτων είναι ελάχιστες, διότι γίνεται λανθασμένη εκτίμηση κόστους κι αυτό έχει ως αποτέλεσμα την αδυναμία της ορθής λειτουργίας τους. Ακόμη, οι πολλοί παράγοντες που έχουν να αντιμετωπίσουν αυτά τα συστήματα καθιστά δυσκολότερη την αποδοτικότητα και την αποτελεσματικότητά τους.

2.6 Ο ΡΟΛΟΣ ΤΗΣ ΕΠΙΧΕΙΡΗΜΑΤΙΚΗΣ ΕΥΦΥΪΑΣ ΣΤΗ ΛΗΨΗ ΕΠΙΧΕΙΡΗΜΑΤΙΚΩΝ ΑΠΟΦΑΣΕΩΝ

Η λήψη των επιχειρηματικών αποφάσεων αποτελεί από τις πιο σημαντικές και πολύπλοκες διαδικασίες σε έναν οργανισμό, διότι διαμορφώνεται σύμφωνα με διάφορες διαστάσεις, οι οποίες επιλέγονται λαμβάνοντας υπόψη το εσωτερικό, αλλά και το εξωτερικό περιβάλλον του. Οι υπεύθυνοι για τις αποφάσεις, που συνήθως είναι μέλη της διοίκησης, καθορίζουν τον τύπο των αποφάσεων που έχουν να διαχειριστούν κι αυτό έχει ως αποτέλεσμα τη δημιουργία του περιβάλλοντος απόφασης. Αυτό βέβαια τροποποιείται λόγω των συνεχών-απρόσμενων αλλαγών στους σύγχρονους ρυθμούς του επιχειρηματικού τομέα και καθιστά τη διαδικασία της απόφασης ακόμα πιο πολύπλοκη.

Η επιχειρηματική ευφυΐα μπορεί να παίξει κομβικό ρόλο στη διαδικασία απόφασης, εφόσον συνδυαστεί και ενσωματωθεί κατάλληλα σε αυτήν. Ο προσδιορισμός των τύπων των αποφάσεων, όπου περιλαμβάνεται στη διαδικασία, καθορίζει τις πληροφορίες και γενικότερα τα δεδομένα που είναι απαραίτητα για την πραγματοποίησή της. Οι αποφάσεις κατατάσσονται σε συνάρτηση με το περιβάλλον απόφασης κι έτσι ο βαθμός στον οποίο είναι δομημένες ή μη-δομημένες επηρεάζει την επιλογή και την απόδοση των αναλυτικών μεθόδων στη διαδικασία της λήψης αποφάσεων, διότι σε αυτό το σημείο καθορίζονται οι απαιτήσεις για την πληροφόρηση των εν λόγω αποφάσεων.

Τα Συστήματα Επιχειρηματικής Ευφυΐας συνήθως υιοθετούνται για την βελτίωση της απόδοσης ήδη εγκατεστημένων πληροφοριακών συστημάτων, όπως ERP, διότι εξελίσσουν τη διαχείριση και την ανάλυση του όγκου των επιχειρηματικών δεδομένων που ήδη διατηρούν. Ωστόσο, οι οργανισμοί για να αντλήσουν αυτά τα πλεονεκτήματα και να επιτύχουν πρέπει η ενσωμάτωση αυτών των συστημάτων να είναι αποτελεσματική σύμφωνα με τις διαδικασίες τους και τη λειτουργία τους. Επιπλέον, η στρατηγική που ακολουθεί ένας οργανισμός μπορεί να επηρεάσει την απόδοση ενός Συστήματος Επιχειρηματικής Ευφυΐας, διότι οι στόχοι του οργανισμού οδηγούν σε δημιουργία αποφάσεων που θα εξεταστούν-αναλυθούν με την υποστήριξη του συστήματος, γι' αυτό είναι αναγκαία η ορθή ενσωμάτωσή του.

Ωστόσο, πολλοί οργανισμοί εκμεταλλεύονται τις δυνατότητες των Συστημάτων Επιχειρηματικής Ευφυΐας χρησιμοποιώντας τα εργαλεία τους όχι μόνο στη

διαδικασία της λήψης επιχειρηματικών αποφάσεων, αλλά ευρύτερα στις επιχειρηματικές δραστηριότητες [13]. Έρευνες όμως έχουν δείξει ότι η πρόοδος-απόδοση αυτών των διαδικασιών και γενικότερα του οργανισμού εξαρτώνται από τη σχέση που τις συνδέει μεταξύ τους [14]. Εξετάστηκε η αξία της χρήσης της επιχειρηματικής ευφυΐας σε δύο επίπεδα, στην απόδοση των επιχειρηματικών διαδικασιών και στην οργανωτική απόδοση του οργανισμού. Τα αποτελέσματα έδειξαν ότι η αξία διαφοροποιείται ανάλογα με τα κριτήρια που μετράται. Για παράδειγμα, η εφαρμογή της επιχειρηματικής ευφυΐας στη διαδικασία παραγωγής μπορεί να φέρει θετικά αποτελέσματα, όπως τη μείωση του κόστους της μονάδας ή τη βελτίωση της παραγωγικής δραστηριότητας, ενώ κατά τη λειτουργία μιας διαδικασίας της αλυσίδας αξίας εξαρτάται κι από άλλους παράγοντες, όπως το ανταγωνιστικό περιβάλλον ή τη στρατηγική που υποστηρίζει ένας οργανισμός. Θεωρείται πως η σχέση που τις συνδέει μεταξύ τους, δηλαδή μια εσωτερική διαδικασία μιας οργάνωσης μπορεί να βελτιωθεί, αλλά να μην εξελίξει απαραίτητα την οργάνωση ευρέως, διότι ο βαθμός της διαδικασίας δεν καθιστά αυτή τη σχέση στενά συνδεδεμένη.

Ακόμη, ένας παράγοντας που θα μπορούσε να επηρεάσει την απόδοση μιας τεχνολογικής λύσης που βασίζεται στη χρήση δεδομένων είναι η στρατηγική συλλογής και διατήρησης αυτών που ακολουθεί ένας οργανισμός. Υπάρχουν δύο είδη αυτής της στρατηγικής και είναι (α) η συλλογική και (β) η εστιασμένη σε συγκεκριμένο στόχο [15]. Η πρώτη δεν ενδεικνύεται, διότι η συλλογή δεδομένων που αφορούν κάθε τμήμα ενός οργανισμού, η ενσωμάτωση αυτών και η διατήρησή τους καθιστά δύσκολη κάθε στοχευόμενη διαδικασία που αφορά την επίλυση ενός συγκεκριμένου προβλήματος. Επιπροσθέτως, ο καθορισμός των στόχων και των προβλημάτων που πρέπει να αντιμετωπίσει μια οργάνωση πραγματοποιούνται από αυτήν σε αρχικό στάδιο, κάτι που είναι δύσκολο και χρονοβόρο όταν έχει να διαχειριστεί έναν ενιαίο όγκο πληροφοριών.

Από την άλλη πλευρά, όταν η συλλογή των δεδομένων υλοποιείται έπειτα από τον προσδιορισμό των σκοπών, που κρίνουν οι υπεύθυνοι της απόφασης, δηλαδή είναι συγκεκριμένη για την εκπλήρωση ενός αιτήματος, δίνεται άμεσα η δυνατότητα της διαχείρισης πιο ενοποιημένων δεδομένων κι αυτό έχει ως αποτέλεσμα μια καλύτερη προσέγγιση για την επίλυση ενός προβλήματος και κυρίως γρηγορότερη. Ακόμη, όταν οι πληροφορίες είναι πιο ομαδοποιημένες για ένα συγκεκριμένο σκοπό αποκτούν μεγαλύτερη αξία, κάτι που έρευνες έχουν δείξει ότι η ποιότητα των δεδομένων συνδέεται θετικά με τη διαχείριση και την πρόοδο των διαδικασιών ενός συστήματος, όπως η διαχείριση ολόκληρων αλυσίδων εφοδιασμού [16] και η ακριβέστερη λήψη απόφασης όχι μόνο σε πραγματικό χρόνο, αλλά και μελλοντικά βάσει πρόβλεψης [17].

Τέλος, οι ταχείς ρυθμοί των εξελίξεων στο επιχειρηματικό περιβάλλον καθιστά αναγκαία τη διαχείριση αβεβαιοτήτων και ρίσκων από τους οργανισμούς κι αποτελεί βασική αιτία επένδυσης σε συστήματα τεχνολογίας που υποστηρίζουν και βελτιώνουν την επιχειρηματική δραστηριότητα. Τα συστήματα επιχειρηματικής ευφυΐας μπορούν να χρησιμοποιηθούν για να δώσουν τις απαντήσεις σε διάφορα τμήματα μιας οργάνωσης, για παράδειγμα στο τμήμα πωλήσεων και μάρκετινγκ. Ερωτήματα όπως, ποιοι είναι οι πιο συχνοί πελάτες, ποιο προϊόν έχει μεγαλύτερες πωλήσεις και ποιοι παράμετροι έχουν επιρροή σε αυτές, τι προτιμάται από το πελατειακό κοινό και γιατί ένα προϊόν παρουσιάζει πτώση στις πωλήσεις του μπορούν να απαντηθούν. Έτσι, αδιαμφισβήτητα μπορούμε να πούμε πως η συμβολή τέτοιων συστημάτων σε έναν

οργανισμό παίζει καθοριστικό ρόλο στη διαδικασία της λήψης αποφάσεων, εφόσον ληφθούν υπόψη κάποιοι σημαντικοί παράγοντες.

2.7 ΣΥΣΤΗΜΑΤΑ ΕΠΙΧΕΙΡΗΜΑΤΙΚΗΣ ΕΥΦΥΙΑΣ ΚΑΙ ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ (BI AND MIS)

Τα συστήματα επιχειρηματικής ευφυΐας διαφέρουν από τα Πληροφοριακά Συστήματα Διοίκησης, δηλαδή τα DSS, τα EIS και τα ES, στην ευρύτερη θεματική τους εμβέλεια. Η πολυπαραγοντική ανάλυση που πραγματοποιούν, τα ημιδομημένα δεδομένα που χρησιμοποιούν, τα οποία προέρχονται από διαφορετικές πηγές και η πολυδιάστατη παρουσίαση αυτών των στοιχείων διαφοροποιούνται μεταξύ αυτών των συστημάτων.

Τα συστήματα BI προσφέρουν μια πρόταση για την αντιμετώπιση των αναγκών των σύγχρονων οργανισμών. Κύρια ζητήματα που πρέπει να χειριστούν αποτελούν η έξυπνη εξερεύνηση, η ολοκλήρωση, η συνάθροιση και η πολυδιάστατη ανάλυση των δεδομένων που προέρχονται από διάφορους πληροφοριακούς πόρους. Στόχος τους αποτελεί ο συνδυασμός δεδομένων από εσωτερικά συστήματα πληροφοριών μιας επιχείρησης και η ενσωμάτωση των δεδομένων που προέρχονται από συγκεκριμένα περιβάλλοντα, όπως στατιστικές, οικονομικές και επενδυτικές πύλες και διάφορες βάσεις δεδομένων.

Στον επιχειρησιακό κόσμο όμως, η δομή και η πολυπλοκότητα των οργανισμών δεν είναι πανομοιότυπες, πράγμα που επηρεάζει τις μεθόδους που ακολουθούν τα πληροφοριακά συστήματα. Κάθε οργανισμός διαμορφώνει τη δική του δομή πέρα από τα επίπεδα που αφορούν τη διοίκησή του. Τα συστήματα BI μπορούν να υποστηρίξουν τη λήψη αποφάσεων σε όλα τα επίπεδα διοίκησης ανεξάρτητα από το επίπεδο της διαρθρωτικοποίησής τους. Δίνουν τη δυνατότητα εκτέλεσης διαφορετικών συγκριτικών αναφορών, για παράδειγμα σε ιστορικά αποτελέσματα, στην κερδοφορία συγκεκριμένων προσφορών, στην αποτελεσματικότητα των διαύλων διανομής καθώς και στην πραγματοποίηση προσομοιώσεων ανάπτυξης ή πρόβλεψης μελλοντικών αποτελεσμάτων έχοντας ως βάση ορισμένες υποθέσεις και παραδοχές. Παρέχει δηλαδή τη δυνατότητα του ακριβή ορισμού των στόχων και την υλοποίηση τους σε στρατηγικό επίπεδο.

Από την πλευρά της μεθοδικότητας, τα συστήματα BI μπορούν να διατηρούν κάποια βάση για τη λήψη αποφάσεων σε θέματα που αφορούν το μάρκετινγκ, τις πωλήσεις, τη χρηματοδότηση, τη διαχείριση κεφαλαίων κ.ά. Η διαμόρφωση ενός πλαισίου εξασφαλίζει στα συστήματα κατά την εφαρμογή τους να επιτυγχάνουν τη βελτιστοποίηση των μελλοντικών ενεργειών, την κατάλληλη τροποποίησή τους με σκοπό την πιο αποτελεσματική υλοποίηση των στόχων.

Ακόμη, σε επιχειρησιακό επίπεδο δίνουν τις απαντήσεις σε ερωτήματα που αφορούν τις τρέχουσες λειτουργίες των τμημάτων ενός οργανισμού και την ενημέρωση σφαιρικά όχι μόνο του εσωτερικού περιβάλλοντός του, αλλά για παράδειγμα και τις εξελίξεις των συνεργασιών του με τους προμηθευτές-πελάτες του.

ΚΕΦΑΛΑΙΟ 3:

POWER BI

3.1 ΕΙΣΑΓΩΓΗ

Το **Power BI** είναι μια σουίτα εργαλείων επιχειρηματικής ανάλυσης της Microsoft που διανέμει πληροφορίες σε ολόκληρο τον οργανισμό. Συνδέεται με εκατοντάδες πηγές δεδομένων, απλοποιεί την προετοιμασία των δεδομένων και πραγματοποιεί ad hoc ανάλυση. Παρέχει τη δυνατότητα δημιουργίας όμορφων αναφορών, οι οποίες μπορούν στη συνέχεια να είναι διαθέσιμες-προσβάσιμες απ' όλο τον οργανισμό, είτε διαδικτυακά είτε μέσω κινητών συσκευών.

3.2 ΓΙΑΤΙ ΧΡΗΣΙΜΟΠΟΙΕΙΤΑΙ

Ο σκοπός της χρήσης αυτής της εφαρμογής χωρίζεται σε τρία επίπεδα, τα οποία παρουσιάζονται παρακάτω.

Εισαγωγή Δεδομένων

Σε αυτό το πρώτο επίπεδο είναι εφικτή η εισαγωγή δεδομένων από μεγάλη ποικιλία πηγών. Υποστηρίζει εταιρικές βάσεις δεδομένων σε αρχεία επιφάνειας εργασίας και δεδομένα σε κοινωνικά δίκτυα. Ακόμη, μπορεί να γίνει συγχώνευση δεδομένων από πολλαπλές πηγές και να τροποποιηθεί ώστε να έχει τη μορφή μιας ενιαίας δομής. Κατά την τροποποίηση αυτή περιλαμβάνεται ο καθαρισμός από μη αξιόπιστα δεδομένα, εστιάζοντας σε εύχρηστα, τα οποία αναλύονται σύμφωνα με τις απαιτήσεις του ενδιαφερόμενου.

Μοντελοποίηση Δεδομένων

Εφόσον έχει πραγματοποιηθεί στο πρώτο στάδιο μια δομή δεδομένων από διάφορες πηγές, σε αυτό το σημείο μπορεί να δημιουργηθεί ένα μοντέλο δεδομένων συνδυάζοντας τους πίνακες που αναπτύχθηκαν κατά την τροποποίηση, ώστε να γίνει η ανάπτυξη μιας συνεκτικής δομής. Επιπλέον, παρέχεται η δυνατότητα μορφοποίησης αυτού του μοντέλου, όπως η επιλογή της εφαρμογής συνεκτικών ονομάτων και τύπων δεδομένων. Έτσι, ο χρήστης προετοιμάζει τις βασικές μετρήσεις που θα χρησιμοποιήσει στις αναλύσεις και τις παρουσιάσεις του στο 3ο επίπεδο.

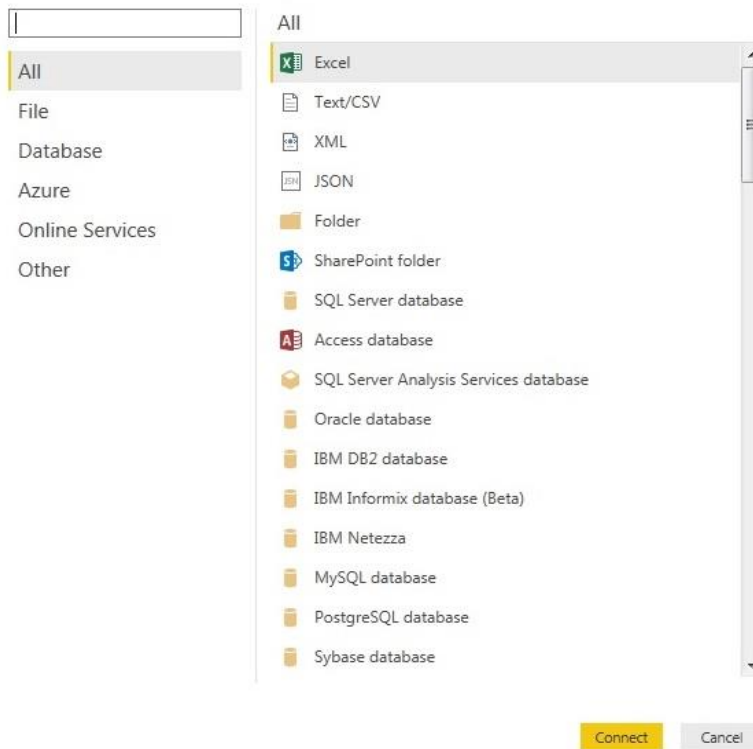
Δημιουργία Αναφορών και Πινάκων Ελέγχου

Στο τελικό αυτό στάδιο παρέχονται πολλές επιλογές παρουσιάσεων των μετρήσεων, όπως πίνακες, διαγράμματα, χάρτες κι άλλα που θα παρουσιαστούν παρακάτω σύμφωνα με ένα σενάριο κάποιων δεδομένων. Υπάρχουν πολλά είδη απεικόνισης που υποστηρίζει το Power BI, όπου η επιλογή τους και η τροποποίησή τους εξαρτάται από τα δεδομένα και την κρίση του χρήστη σύμφωνα με την εξαγωγή συμπερασμάτων-απαντήσεων στα αιτήματά που τον ενδιαφέρουν.

3.3 ΤΟ ΠΕΡΙΒΑΛΛΟΝ ΤΟΥ POWER BI DESKTOP

Για την παρουσίαση του περιβάλλοντος της εφαρμογής του Power BI θα υλοποιήσουμε ένα σενάριο στο Power BI Desktop, το οποίο αποτελεί εργαλείο για τον υπολογιστή και μας επιτρέπει να δημιουργήσουμε μια συλλογή ερωτημάτων, συνδέσεων δεδομένων και αναφορών που μπορούν να είναι προσβάσιμα κι από άλλους χρήστες.

Get Data



Με την εφαρμογή του θα εξετάσουμε τις δυνατότητες που προσφέρει και τις επιλογές που μας παρέχει ευρύτερα. Όπως βλέπουμε στην Εικόνα 3.1, υπάρχουν πολλές επιλογές μορφής εισαγωγής δεδομένων και όπως έχουμε αναφέρει παραπάνω είναι εφικτή και η εισαγωγή διαφορετικών τύπων πηγών δεδομένων. Ορισμένες πηγές δεδομένων έχουν ήδη τα δεδομένα τους σε κατάλληλη μορφή για τις υπηρεσίες του Power BI, όπως πακέτα από παρόχους υπηρεσιών (π.χ. Google analytics).

Εικόνα 3.1: Εισαγωγή Δεδομένων

Σε άλλες περιπτώσεις, ίσως χρειαστεί να πραγματοποιήσουμε κάποια αναζήτηση για να συλλέξουμε τα δεδομένα που μας ενδιαφέρουν και να τα τοποθετήσουμε σε ένα αρχείο, το οποίο θα χρησιμοποιήσετε στο πρόγραμμα.

Στην υπηρεσία του Power BI, δε μπορούμε να συνδεθούμε απευθείας με μία βάση δεδομένων και να ξεκινήσουμε την εξερεύνηση των δεδομένων της εάν δεν είναι δομημένη με πίνακες μοντέλων. Ωστόσο, με την υπηρεσία του Power BI Desktop ή το Excel μπορούμε να προ-επεξεργαστούμε τα δεδομένα μας ώστε να έχουν την κατάλληλη ή την επιθυμητή μορφή για την μετέπειτα ανάλυσή τους χρησιμοποιώντας τα εργαλεία που διαθέτει αυτό το πρόγραμμα.

Όταν τα δεδομένα θεωρούμε ότι είναι καθαρά κι έτοιμα για χρήση στα αρχεία που έχουμε, τα επιλέγουμε και προχωράμε στη φόρτωσή τους στο πρόγραμμα. Τη στιγμή που θα επιλέξουμε τη φόρτωση το πρόγραμμα διαβάζει τα δεδομένα κι ένα σύνολο δεδομένων (data set) δημιουργείται αυτόματα στο Power BI, το οποίο περιέχει πληροφορίες σχετικά με τις πηγές των δεδομένων και σχεδόν πάντα ένα υποσύνολο των δεδομένων όπου αντιγράφονται από τις πηγές δεδομένων. Αφού ολοκληρωθεί η φόρτωση των αρχείων με τα δεδομένα, το πρόγραμμα εμφανίζει κάποιες επιλογές με βάση τους ορισμένους πίνακες, τα φύλλα εργασίας ή τις ονομαστικές περιοχές των δεδομένων. Αυτό συμβαίνει διότι πολλά αρχεία περιλαμβάνουν πλήθος πινάκων. Σε αυτό το σημείο διαθέτεται η προεπισκόπηση της μορφής των συνόλων των δεδομένων που δημιουργήθηκαν και είναι προτιμότερο να επιλέγουμε με γνώμονα

τους τίτλους των πινάκων, γιατί σε ένα "συνολικό" σύνολο δεδομένων (π.χ. Φύλλο εργασίας) υπάρχει μεγάλη πιθανότητα να υπάρχουν κενά ή δεδομένα εκτός των πινάκων, τα οποία θα δημιουργήσουν δυσκολία στη διαδικασία της φόρτωσης.

Επιπλέον, αφού επιλέξουμε τα σύνολα των δεδομένων που δημιουργήθηκαν, το πρόγραμμα μας δίνει τη δυνατότητα να τα επεξεργαστούμε με το εργαλείο Query Editor πριν τη φόρτωσή τους. Αυτό το εργαλείο διαθέτει αρκετές δυνατότητες πάνω στα δεδομένα. Αν επιθυμούμε κάποια αλλαγή σε κάποιον πίνακα, όπως αύξουσα ή φθίνουσα ταξινόμηση, διαγραφή κελιών, καθαρισμό από διπλότυπα και εφαρμογή φίλτρων είτε σε ονομαστικές είτε σε αριθμητικές μεταβλητές. Ακόμη, μπορούμε να μορφοποιήσουμε ή να τροποποιήσουμε τους τύπους των δεδομένων. Τέλος, οποιαδήποτε προσθήκη νέων δεδομένων είναι εφικτή να πραγματοποιηθεί με αυτό το εργαλείο. Για παράδειγμα, η δημιουργία κάποιων καινούριων πινάκων δεδομένων.

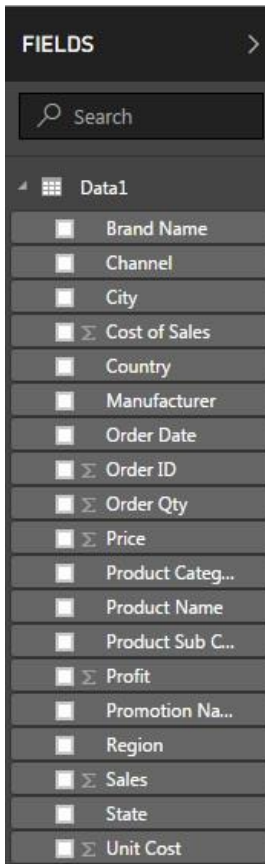
Στο δικό μας σενάριο η πηγή μας θα είναι ένα αρχείο Excel που θα περιλαμβάνει δεδομένα που αφορούν παραγγελίες προϊόντων τεχνολογίας. Οι στήλες του περιλαμβάνουν στοιχεία που αφορούν με τη σειρά τα εξής:

1. Order ID
2. Order Date
3. Unit Cost
4. Price
5. Order Qty
6. Cost of Sales
7. Sales
8. Profit
9. Channel
10. Promotion Name
11. Product Name
12. Manufacturer
13. Brand Name
14. Product Sub Category
15. Product Category
16. Region
17. City
18. State
19. Country

Για να γίνει πιο κατανοητό ένα παράδειγμα είναι η ακόλουθη γραμμή του αρχείου:

7077 | 18/3/2013 | 76,09496775 | 304 | 9 | 684,8547097 | 2714,72 | 2029,86529 |
Store | European Spring Promotion | Contoso SLR Camera M143 Grey | Contoso, Ltd
| Contoso | Digital SLR Cameras | Cameras and camcorders | Europe | Moscow |
Moskovskaya oblast | Russia

Στη συνέχεια, αφού επιλέξουμε το αρχείο Excel από όπου θα προμηθευτεί τα δεδομένα, δημιουργεί τα πεδία τα οποία θα χρησιμοποιήσουμε στα είδη απεικόνισης που παρέχει το πρόγραμμα. Όπως φαίνεται στην Εικόνα 3.2, βλέπουμε τα πεδία που δημιουργήθηκαν και σε αυτό το σημείο είναι σημαντικό να αναφέρουμε ότι είναι εφικτή η τροποποίησή τους και η ανάπτυξή τους γενικότερα.

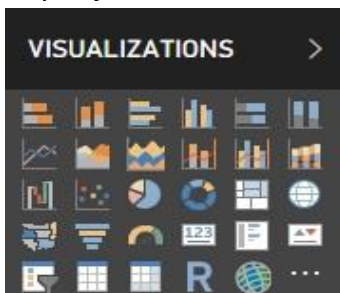


Στο δικό μας σενάριο, υπάρχουν στήλες με αριθμητικά στοιχεία, τα οποία έχουν πολλά ψηφία, όπως φαίνεται στο παράδειγμα του αρχείου παραπάνω, κάτι που θα καθιστούσε δυσκολότερη την ερμηνεία μιας απεικόνισης. Έτσι, λόγω της δυνατότητας της τροποποίησης μπορεί να διαμορφωθεί όπως επιθυμούμε. Ακόμη, παρατηρώντας τα πεδία δημιουργούνται πολλά υποθετικά ερωτήματα, είτε για τα ύψη των πωλήσεων είτε για τις παραγγελίες, και τα περισσότερα σε συνάρτηση με το χρόνο. Οπότε, πέρα από το πεδίο Order Date, θεωρείται απαραίτητη η δημιουργία νέων πεδίων με βάση το χρόνο, τα οποία θα είναι το Year και το Month. Έτσι, πηγαίνουμε στην διεπαφή με τα δεδομένα, όπου διαθέτει ανασκόπηση και μοντελοποίηση των δεδομένων και με τη χρήση των εκφράσεων ανάλυσης δεδομένων (Data Analysis Expressions) που μας παρέχει το πρόγραμμα, δημιουργούμε δύο νέες στήλες στον πίνακα, την Year και την Month, αντλώντας από το πεδίο Order Date το έτος (Year=Year([Order Date])) και τον μήνα (Month=Month([Order Date])).

Εικόνα3.2: Πεδία Δεδομένων

Οι εκφράσεις ανάλυσης δεδομένων ή DAX είναι συλλογές από συναρτήσεις με τις οποίες μπορούμε υπολογίσουμε την επιστροφή της τιμής ή των τιμών σε ένα τύπο δεδομένων. Ακόμη, μας βοηθούν να αντλήσουμε παραπάνω πληροφορίες από τα δεδομένα μας, γεγονός που θα δούμε και παρακάτω στη δημιουργία των αναφορών μας. Αυτές οι διαδικασίες δηλαδή, μπορούν να υλοποιηθούν οποιαδήποτε στιγμή, αλλά είναι σημαντικό να σημειωθεί πως ο χρόνος αποτελεί κύριο και αναπόσπαστο κομμάτι μιας αναφοράς-απεικόνισης.

Αρχικά θα δημιουργήσουμε απεικονίσεις με τη μορφή πινάκων επιλέγοντας κάποια από τα πεδία με σκοπό να απαντήσουμε σε ερωτήματα όπως "Ποιος κατασκευαστής έχει τις υψηλότερες πωλήσεις;", "Ποια κατηγορία προϊόντος έχει το χαμηλότερο κέρδος;", "Ποιο κανάλι έχει το μεγαλύτερο κόστος πωλήσεων;", "Ποιο όνομα προώθησης έχει τις υψηλότερες πωλήσεις;", "Ποια υπό-κατηγορία προϊόντος έχει το μεγαλύτερο κέρδος;".



Όπως βλέπουμε στην Εικόνα 3.3, μπορούμε να απεικονίσουμε τα δεδομένα μας με διάφορες μορφές. Μερικές από αυτές είναι οι πίνακες, τα ραβδογράμματα, τα ιστογράμματα, τα διαγράμματα πίτας, τα διαγράμματα διασποράς, γεωγραφικά βάσει χαρτών και μέσω καρτών πολλαπλών ροών.

Εικόνα 3.3: Είδη Απεικόνισης

Ακόμη, παρέχεται η δυνατότητα παρέμβασης στη μορφή απεικόνισης χρησιμοποιώντας τη γλώσσα προγραμματισμού R.

Ας δούμε όμως κάποια παραδείγματα πινάκων που δίνουν ερμηνεία στα ερωτήματα που αναφέρθηκαν προηγουμένως.

Για να δούμε για παράδειγμα ποιος κατασκευαστής έχει τις υψηλότερες πωλήσεις, θα επιλέξουμε το εικονίδιο με τον απλό πίνακα και ως παραμέτρους του τα δύο πεδία, τα οποία είναι Manufacturer και Sales (βλ. Εικόνα 3.4). Ακόμη, για το ποια κατηγορία προϊόντος έχει το χαμηλότερο κέρδος, τα πεδία θα είναι Profit και Product Category (βλ. Εικόνα 3.5). Όπως παρατηρούμε στις εικόνες ο κατασκευαστής με τις πιο πολλές πωλήσεις είναι ο Fabrikam, Inc. και η κατηγορία προϊόντων Audio έχει το χαμηλότερο κέρδος. Θα μπορούσαμε να προσθέσουμε κι άλλα πεδία, όπως και κάναμε σε αυτές τις δύο περιπτώσεις για να δείξουμε πως με την επιλογή της ταξινόμησης κάποιας στήλης αναδιαμορφώνεται ο πίνακας κι έτσι εξάγουμε κι άλλα συμπεράσματα, όπως τι κέρδος λαμβάνουμε σύμφωνα με τα ύψη των πωλήσεων ανά κατασκευαστή.

Manufacturer	Sales	Profit
Fabrikam, Inc.	12.054.281	7158977
Contoso, Ltd	9.996.750	5670160
Adventure Works	7.467.704	4158548
Proseware, Inc.	5.996.829	3513125
Wide World Importers	5.700.910	3198143
A. Datum Corporation	4.404.982	2608508
The Phone Company	4.386.164	2476148
Southridge Video	3.112.730	1610013
Litware, Inc.	2.134.837	1120287
Northwind Traders	136.572	73528
Total	55.391.760	31587437

Εικόνα 3.4

Sales	Profit	Product Category
1.029.381	577270	Audio
1.057.480	635285	Music, Movies and Audio Books
5.841.156	3285926	Cell phones
9.107.839	4823735	TV and Video
17.048.932	10102802	Cameras and camcorders
21.306.971	12162419	Computers
55.391.760	31587437	

Εικόνα 3.5

Ένας άλλος τρόπος απεικόνισης είναι η χρήση matrix πίνακα, όπου μπορούμε να συνδυάσουμε περισσότερα πεδία, τα οποία μπορούν να ερμηνευτούν καλύτερα. Οι πίνακες αυτοί διευκολύνουν πολύ τους αναγνώστες, διότι οργανώνουν τις πληροφορίες ανάλογα με τα πεδία που τους προμηθεύουμε. Ένα τμήμα τέτοιου πίνακα παρουσιάζεται στην Εικόνα 3.6, όπου απεικονίζει τις πωλήσεις για κάθε κατασκευαστή σε όλα τα κανάλια (Catalog, Online, Reseller, Store) σε κάθε περιοχή (Asia, Europe, North America). Κατά τη δημιουργία ενός πίνακα matrix θα μπορούσαμε να λάβουμε υπόψη τις ιεραρχίες που υπάρχουν ή έχουμε δημιουργήσει στα δεδομένα μας.

Manufacturer	Asia	Europe	North America	Total
A. Datum Corporation	1.095.334	787.911	2.521.737	4.404.982
Catalog			366.077	366.077
Online	423.496	304.896	308.348	1.036.740
Reseller	242.590	122.112	189.628	554.331
Store	429.248	360.902	1.657.684	2.447.834
Adventure Works	1.641.526	1.468.547	4.357.630	7.467.704
Catalog			596.047	596.047
Online	433.489	559.792	641.124	1.634.405
Reseller	355.006	236.694	309.221	900.921
Store	853.031	672.062	2.811.238	4.336.331
Contoso, Ltd	2.189.425	1.833.012	5.974.313	9.996.750
Catalog			1.014.299	1.014.299
Online	754.574	612.568	889.344	2.256.486
Reseller	242.590	122.112	189.628	554.331
Store	429.248	360.902	1.657.684	2.447.834
Total	12.412.345	10.195.858	32.783.557	55.391.760

Εικόνα 3.6

Στο παράδειγμά μας βλέπουμε ότι υπάρχει μία λογική σειρά, δηλαδή πρώτα πάει ο κατασκευαστής, δεύτερα τα κανάλια διανομής και τέλος οι περιοχές. Ακολουθώντας δηλαδή κάποια επίπεδα ιεραρχίας σύμφωνα πάντα με τα δεδομένα μας, οι πίνακες matrix μπορούν να μας προσφέρουν πολλές πληροφορίες σε αρκετά πεδία παράλληλα, διότι ομαδοποιούν αυτόματα κάθε στοιχείο με την κατηγορία του. Όπως και στους προηγούμενους πίνακες έτσι και σε αυτόν ισχύει η αυτόματη επιλογή της ταξινόμησης. Με ένα κλικ σε κάποιο από τα πεδία πάνω στον πίνακα θα πραγματοποιηθεί ταξινόμηση και στην περίπτωση αυτού του πίνακα θα ταξινομηθούν ανάλογα και οι ιεραρχίες, δηλαδή τα υπόλοιπα πεδία του πίνακα. Αυτή η δυνατότητα σε συνδυασμό με αυτό τον τύπο πίνακα μας διαθέτει μια πιο στοχευόμενη αναζήτηση πληροφορίας ανάμεσα σε πολλά πεδία. Συνήθως αυτοί οι πίνακες είναι μεγάλοι, αλλά η ευκολία στην ερμηνεία τους από τους αναγνώστες τους κάνει να ξεχωρίζουν.

Ακόμη μια επιλογή απεικόνισης που διαθέτει το πρόγραμμα είναι αυτή με τα ομαδοποιημένα διαγράμματα ράβδων (clustered bar chart) και τα ομαδοποιημένα διαγράμματα στηλών (clustered column chart).

Ένα παράδειγμα στο δικό μας σενάριο θα ήταν να δούμε τα ύψη των πωλήσεων, του κόστους των πωλήσεων και του κέρδους σε όλα τα κανάλια διανομής. Όπως φαίνεται στην Εικόνα 3.7.



Εικόνα 3.7

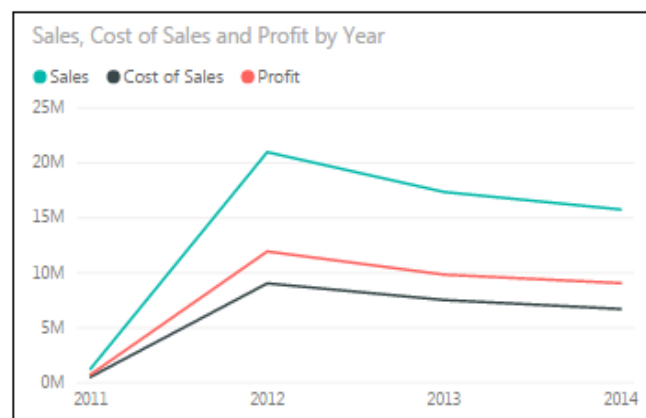
Οι υψηλότερες πωλήσεις πραγματοποιούνται στο κανάλι διανομής Store ενώ οι αμέσως επόμενες είναι οι Online πωλήσεις. Ακόμη, παρατηρούμε ότι τα κόστη των πωλήσεων (Cost of Sales) είναι περίπου στο μισό του ύψους των πωλήσεων (Sales). Ένα σημείο άξιο παρατήρησης είναι ότι το κέρδος (Profit) είναι υψηλότερο από το κόστος των πωλήσεων σε κάθε κανάλι. Αυτό σύμφωνα με τα δεδομένα του σεναρίου μας δείχνει μια καλή εικόνα του οργανισμού. Βέβαια, με τα διαγράμματα ράβδων και στηλών δε μπορούμε να αντλήσουμε πληροφορίες με μεγάλη λεπτομέρεια, αλλά μπορούμε να αντιληφθούμε σημαντικές διαφορές σε συγκρίσιμα πεδία ή πεδία που χρειάζονται περαιτέρω εξέταση κι ανάλυση.

Μία απεικόνιση με ομαδοποιημένο διάγραμμα στηλών παρουσιάζεται στην Εικόνα 3.8, όπου τα πεδία που επιλέξαμε είναι οι πωλήσεις, τα κόστη τα πωλήσεων, το κέρδος και η περιοχή (Region). Έτσι, βλέπουμε ότι στη Βόρεια Αμερική πραγματοποιούνται με διαφορά οι υψηλότερες πωλήσεις. Αυτό θα μπορούσε να μας οδηγήσει σε παραπάνω έρευνα πάνω στα προϊόντα που πωλούνται στην Βόρεια Αμερική.



Εικόνα 3.8

Όπως προαναφέρθηκε το πεδίο του χρόνου είναι από τα πιο σημαντικά πεδία στην ανάλυση των δεδομένων, διότι μπορούμε να συγκρίνουμε και να εξετάσουμε την εξέλιξη, αν όχι όλων, αρκετών μεταβλητών. Η Εικόνα 3.9 απεικονίζει ένα διάγραμμα γραμμής (Line chart), το οποίο παρουσιάζει τις πωλήσεις, τα κόστη των πωλήσεων και το κέρδος από το 2011 έως το 2014. Το πεδίο Year είναι το πεδίο που δημιουργήσαμε στην αρχή από τα δεδομένα μας. Έτσι, μπορούμε να δούμε ανά έτος την πορεία αυτών των τριών και άλλων μεταβλητών.



Εικόνα 3.9

Ακόμη μία μορφή απεικόνισης είναι η κάρτα πολλαπλών γραμμών (Multi-row card), η οποία εμφανίζει τα επιλεγμένα πεδία σε γραμμές. Η απεικόνιση των πινάκων με αυτή τη μορφή είναι από τις καλύτερες επιλογές, διότι ανάλογα με το πεδίο που θέλουμε να εξετάσουμε αναπροσαρμόζεται κι αλλάζει μορφή. Έστω ότι έχουμε διάφορους πίνακες, όπως αυτούς που δημιουργήσαμε και δείξαμε παραπάνω (Εικ.3.4, 3.5, 3.6) και την κάρτα πολλαπλών γραμμών που φαίνεται στην Εικόνα 3.10, η οποία περιλαμβάνει τις πωλήσεις, το κέρδος και την ποσότητα των παραγγελιών (Order Qty). Σε περίπτωση που επιλέξουμε κάποιον κατασκευαστή από τον πίνακα της Εικόνας 3.4, η κάρτα μας θα προσαρμοστεί με βάση την επιλογή μας, δηλαδή θα μπορούμε να δούμε τις πωλήσεις, το κέρδος και την ποσότητα των παραγγελιών σε κάθε περιοχή για τον κατασκευαστή που επιλέξαμε.

Asia		
12.412.345	7144694	57581
Sales	Profit	Order Qty
Europe		
10.195.858	5773913	44592
Sales	Profit	Order Qty
North America		
32.783.557	18668831	148953
Sales	Profit	Order Qty

Εικόνα 3.10

Επίσης, ένας πιο διαδραστικός τρόπος παρουσίασης δεδομένων είναι η κάρτα πολλαπλών ροών με τη χρήση slicer, όπου αφού επιλεγθούν τα πεδία που μας ενδιαφέρουν και τοποθετηθούν στην αναφορά, διαλέγουμε το πεδίο που θα χρησιμοποιήσουμε ως slicer. Στην Εικόνα 3.11 βλέπουμε μία κάρτα με διάφορες ροές, όπως Sales, Region-Sales-Profit (πολλαπλών ροών) και slicer έχουμε επιλέξει την κατηγορία προϊόντος (Product Category), όπου ανάλογα την κατηγορία που διαλέγουμε αναδιαμορφώνεται όλη η αναφορά, δηλαδή στην εικόνα όλα τα νούμερα αντιπροσωπεύουν την κατηγορία προϊόντων Audio. Αποτελεί από τα πιο χρήσιμα εργαλεία, διότι μπορούμε σε πραγματικό χρόνο να εξετάζουμε πολλά πεδία παράλληλα και σύμφωνα με τις προτιμήσεις μας.

1M	577K	1.406	106	32
Sales	Profit	Average of Sales	Count of Product Name	Count of Country
Asia				
253.883	127845			
Sales	Profit			
Europe				
205.473	123204			
Sales	Profit			
North America				
570.025	326221			
Sales	Profit			

Product Category
☐ Select All
☒ Audio
☐ Cameras and camcorders
☐ Cell phones
☐ Computers
☐ Music, Movies and Audio Books
☐ TV and Video

Εικόνα 3.11

Αφού είδαμε κάποιες μορφές απεικόνισης των δεδομένων μας, μπορούμε να δούμε πιο ισχυρές παραμετροποιήσεις που μπορούν να βελτιώσουν-εξελιξουν περισσότερο τα αποτελέσματα και τα συμπεράσματα που μας φέρουν. Μία ακόμη δυνατότητα που διαθέτει το πρόγραμμα αυτό είναι η δημιουργία κι η διαχείριση ιεραρχιών. Αναφερθήκαμε και παραπάνω στην Εικόνα 3.6 στο ενδεχόμενο δημιουργίας ιεραρχιών, αλλά ουσιαστικά θα το αναλύσουμε σε αυτό το σημείο με παράδειγμα στα δεδομένα μας.

Με τον όρο ιεραρχία εννοούμε όταν μια μεταβλητή μας, είτε αριθμητική είτε κατηγορική, μπορεί να χωριστεί σε μία ή περισσότερες κατηγορίες ή σε επίπεδα. Για παράδειγμα, στα δικά μας τα δεδομένα έχουμε πεδία, τα οποία μπορούν να ομαδοποιηθούν ιεραρχικά. Ένα από αυτά είναι η περιοχή (Region). Οπότε πάμε στην καρτέλα με τα πεδία δεδομένων (βλ. Εικόνα 3.2) και δημιουργούμε ιεραρχία με πρώτο επίπεδο το πεδίο Region. Μετά εντάσσουμε τα πεδία που θεωρούμε στα επόμενα επίπεδα της ιεραρχίας. Στη δική μας περίπτωση επιλέγουμε χώρα (Country) και πόλη (City). Αφού ολοκληρώσουμε τις επιλογές μας, δημιουργείτε αυτόματα ένα

ιεραρχημένο πεδίο για την περιοχή (Region Hierarchy), όπου παρουσιάζεται στην Εικόνα 3.12. Την χρησιμότητα αυτής της ιεραρχίας θα την κατανοήσουμε καλύτερα



Εικόνα 3.12

όταν χρησιμοποιήσουμε το πεδίο της ιεραρχίας σε κάποια απεικόνισή μας. Έστω ότι θέλουμε να δούμε το κέρδος μας σε κάθε χώρα για κάθε κατηγορία των προϊόντων μας. Σε άλλη περίπτωση, δηλαδή χωρίς τη δημιουργία της παραπάνω ιεραρχίας θα είχαμε να επιλέξουμε αρκετά πεδία, ενώ τώρα θα δημιουργήσουμε μια απεικόνιση που θα περιλαμβάνει μόνο το πεδίο της ιεραρχίας και τα πεδία που θέλουμε να εξετάσουμε, δηλαδή το κέρδος (Profit) και την κατηγορία προϊόντος (Product Category). Η απεικόνιση που θα χρησιμοποιήσουμε είναι ο matrix πίνακας, διότι αποτελεί ιδανική επιλογή όταν έχουμε να εξετάσουμε αρκετά πεδία. Στην Εικόνα 3.13 βλέπουμε μέρος του πίνακα που δημιουργήσαμε, διότι η υποκατηγορία χώρα της ιεραρχίας περιλαμβάνει πολλές μεταβλητές.

Region	Audio	Cameras and camcorders	Cell phones	Computers	Music, Movies and Audio Books	TV and Video	Total
Asia	127845	2272382	810312	2790546	114765	1028843	7144694
Armenia	1889	45488	12429	27829		24335	111970
Australia	5585	74879	57582	157204	5896	43088	344234
Bhutan	802	40759	7741	30559	363	7951	88175
China	76059	1467743	524526	1704075	70663	549304	4392369
India	6467	92870	45477	149599	7885	55360	357657
Iran	433	39286	8749	116691	3577	23825	192562
Japan	12099	211156	63527	203985	4948	59337	555052
Kyrgyzstan		11866	4813	36116	928	17424	71146
Pakistan	2765	33986	8086	87830	5114	31651	169432
Singapore	602	6040	4188	26189		4813	41833
South Korea	10143	20481	20984	55788		43241	150637
Syria	1465	39922	21731	63532	890	51493	179032
Taiwan	395	45565	3494	31929	1049	26841	109273
Thailand	1180	48124	12108	51850	9091	29468	151821
Turkmenistan	7961	94218	14878	47370	4360	60713	229500
Europe	123204	1787204	581230	2286362	145524	850388	5773913
Denmark	1406	11811	6123	10939	3389	11526	45194
France	20140	417580	168523	550008	60779	184559	1401590
Germany	54961	731680	199976	1016165	50057	420443	2473282
Greece	562	9519	3760	22048	22	7082	42993

Εικόνα 3.13

Στον πίνακα αυτόν, όπως παρατηρούμε μπορούμε να αντλήσουμε αρκετή πληροφορία που αφορά το κέρδος και τις κατηγορίες των προϊόντων. Μας δίνεται η δυνατότητα να δούμε το κέρδος ανά περιοχή και πόσο αντιστοιχεί στις χώρες της περιοχής για κάθε κατηγορία των προϊόντων. Θα μπορούσαμε να περιλαμβάνουμε και την πόλη, αλλά το πρόγραμμα μας διαθέτει την επιλογή των πεδίων και στο πεδίο ιεραρχίας που δημιουργήσαμε.

Ένα άλλο παράδειγμα, έστω ότι θέλαμε να δούμε τις πωλήσεις, το κέρδος και την ποσότητα των παραγγελιών σε κάθε περιοχή, χώρα και πόλη. Με τη δημιουργία ενός matrix πίνακα κι επιλέγοντας τα πεδία που μας ενδιαφέρουν και το πεδίο της ιεραρχίας θα λαμβάναμε τα αποτελέσματα που παρουσιάζονται στην Εικόνα 3.14.

Έτσι μπορούμε να δούμε για κάθε χώρα και πόλη των περιοχών, όπως στο Παρίσι, το ύψος των πωλήσεων μας, το κέρδος μας και τις ποσότητες των παραγγελιών μας σ' εκείνη την πόλη. Αυτή η εικόνα επίσης αποτελεί μέρος του πίνακα, διότι τα πεδία περιλαμβάνουν πολλές κατηγορίες.

Region	Sales	Profit	Order Qty
Europe	10.195.858	5773913	44592
Denmark	79.145	45194	386
Copenhagen	79.145	45194	386
France	2.628.401	1401590	11850
Lyon	60.125	34217	254
Marseille	11.092	4711	95
Nantes	112.748	67957	318
Nice	85.802	37991	780
Paris	2.156.624	1147515	9451
Strasbourg	94.719	46586	320
Toulouse	107.290	62613	632
Germany	4.314.956	2473282	17789
Bamberg	138.774	76826	406
Baumholder	153.510	80682	378
Berlin	3.255.279	1871291	13841
Dusseldorf	71.707	48499	157
Eindhoven	60.194	33073	257
Total	10.195.858	5773913	44592

Εικόνα 3.14

Αφού είδαμε πως με τη χρήση μιας ιεραρχίας μπορούμε να επεκτείνουμε τις παραμέτρους-διαστάσεις μιας απεικόνισης δεδομένων, θα δούμε παρακάτω πώς με τη χρήση φίλτρων και τεχνικών μπορούμε να συγκεκριμενοποιήσουμε τα αποτελέσματά μας.

Κατά τον ίδιο τρόπο, δημιουργούμε μια ιεραρχία για τα προϊόντα μας, η οποία περιλαμβάνει το όνομα του προϊόντος, την κατηγορία του προϊόντος και την υποκατηγορία προϊόντος. Έστω ότι θέλουμε να δούμε ποιο προϊόν φέρει το μεγαλύτερο κέρδος στην περιοχή της Ευρώπης, το ύψος των πωλήσεων του και την ποσότητα των παραγγελιών του. Θα δημιουργήσουμε έναν πίνακα matrix με πεδία τις πωλήσεις, το κέρδος, την ποσότητα των παραγγελιών και το πεδίο της ιεραρχίας που δημιουργήσαμε για το προϊόν. Επίσης, θα δημιουργήσουμε έναν κόφτη (slider) με το πεδίο της ιεραρχίας για την περιοχή και θα του βάλουμε φίλτρο, το οποίο θα κόβει την εμφάνιση των πόλεων και θα μας δίνει την επιλογή μόνο των περιοχών.

Product Name	Sales	Profit	Order Qty
Fabrikam Trendsetter 2/3" 17mm X100 White	78.459	61562	83
Cameras and camcorders	78.459	61562	83
Camcorders	78.459	61562	83
Contoso Projector 1080p X980 Silver	89.620	52123	40
Computers	89.620	52123	40
Projectors & Screens	89.620	52123	40
Fabrikam Laptop15.4W M5480 White	72.010	48976	96
Computers	72.010	48976	96
Laptops	72.010	48976	96
Fabrikam Trendsetter 2/3" 17mm X100 Black	59.800	45761	63
Cameras and camcorders	59.800	45761	63
Camcorders	59.800	45761	63
Proseware Projector 1080p DLP86 Black	54.978	44998	22
Computers	54.978	44998	22
Projectors & Screens	54.978	44998	22
Proseware Projector 720p DLP86 Silver	65.764	43406	87
Total	10.195.858	5773913	44592

Region
☐ Select All
☐ Asia
☒ Europe
☐ North America

Εικόνα 3.15

Τα αποτελέσματα της απεικόνισής μας φαίνονται στην Εικόνα 3.15, όπου έχουμε κάνει ταξινόμηση στη στήλη Profit και το πρώτο προϊόν (Fabrikam Trendsetter 2/3" 17mm X100 White) είναι το προϊόν που μας φέρει το μεγαλύτερο κέρδος στην περιοχή της Ευρώπης. Ακόμη, βλέπουμε στις επόμενες σειρές, την κατηγορία και την

υποκατηγορία του προϊόντος που ανήκει αντίστοιχα. Δεξιά του πίνακα φαίνεται ο κόφτης που δημιουργήσαμε, με τον οποίο μπορούμε να εξετάσουμε και τις άλλες περιοχές.

Αφού είδαμε μερικές μορφές απεικονίσεων, τεχνικών και τη σημασία της δημιουργίας ιεραρχιών στα δεδομένα μας, μπορούμε να αναφέρουμε και πιο ισχυρές παραμετροποιήσεις που να αφορούν τις σχέσεις που υπάρχουν ανάμεσα στα δεδομένα μας. Λέγοντας σχέσεις εννοούμε ότι τα δεδομένα μας είναι δομημένα με τέτοιο τρόπο ώστε η πρόσβαση σε αυτά να πραγματοποιείται μέσω της σχέσης τους με άλλα. Για να γίνει κατανοητό σκεφτείτε πως έχουμε αρκετούς πίνακες στα δεδομένα μας όπου πολλά από τα πεδία τους θα θέλαμε να τα χρησιμοποιήσουμε χωρίς να παίρνουμε ολόκληρους τους πίνακες. Με τη δημιουργία των σχέσεων μεταξύ των πινάκων μας επιτρέπεται η χρήση των στηλών ενός πίνακα σε έναν άλλον. Για παράδειγμα, σκεφτείτε έναν πίνακα που θα αφορούσε αποκλειστικά το προϊόν, δηλαδή τα πεδία που θα περιελάμβανε θα ήταν χαρακτηριστικά μόνο γι' αυτό. Και έναν άλλον πίνακα που θα αφορούσε αποκλειστικά τις παραγγελίες των προϊόντων. Μία σχέση μεταξύ τους θα μπορούσε να δημιουργηθεί με κλειδί τον κωδικό του προϊόντος, έτσι όλα τα πεδία από τον πίνακα του προϊόντος θα μπορούσαν να χρησιμοποιηθούν μέσω του πίνακα των παραγγελιών. Το πρόγραμμα δηλαδή μας διαθέτει τη δυνατότητα δημιουργίας σχέσεων και παραμετροποίηση αυτών όπως επιθυμούμε.

Στη δική μας περίπτωση, τα δεδομένα μας δεν έχουν μεγάλο όγκο ώστε να ομαδοποιηθούν και να δομηθούν τόσο πολύ, αλλά φανταστείτε πως στην πραγματικότητα αυτά τα προγράμματα τροφοδοτούνται από αρκετές και διαφορετικές πηγές δεδομένων, όπου ο σχεδιασμός της δομής αποτελεί σημαντική διαδικασία για την μετέπειτα ανάλυση. Βέβαια, συνήθως τα δεδομένα είναι δομημένα σύμφωνα με το σχεσιακό μοντέλο των βάσεων δεδομένων, αλλά όταν έχουμε πολλαπλά αρχεία από διαφορετικές πηγές το πρόγραμμα τα μοντελοποιεί κατά την φόρτωσή τους σε αυτό.

Η διαδικασία που ακολουθεί το Power BI Desktop κατά τη φόρτωση δεδομένων περιλαμβάνει την προσπάθειά του να μετατρέψει τους τύπους των δεδομένων από τις πηγές με τέτοιο τρόπο ώστε η αποθήκευση, οι υπολογισμοί και η οπτικοποίησή τους να πραγματοποιηθούν αποδοτικότερα. Για παράδειγμα, εάν μια στήλη με αριθμητικές τιμές δεν έχει δεκαδικά ψηφία, το πρόγραμμα θα μετατρέψει ολόκληρη τη στήλη σε έναν τύπο δεδομένων ολόκληρου αριθμού, ο οποίος είναι καταλληλότερος για την αποθήκευση ακέραιων αριθμητικών μεταβλητών. Βέβαια, αυτόν τον τύπο δεδομένων μπορούμε να τον αλλάξουμε στην πορεία αν επιθυμούμε, διότι συνήθως κατά τη δημιουργία αναφορών χρησιμοποιούμε αρκετές φορές τις ίδιες στήλες.

Από την άλλη πλευρά, ο σωστός προσδιορισμός των τύπων των δεδομένων είναι απαραίτητος για ορισμένες DAX (Data Analysis Expressions) λειτουργίες, όπου απαιτούν τα δεδομένα να πληρούν κάποιες προϋποθέσεις. Σε πολλές περιπτώσεις έχει προνοήσει το πρόγραμμα για αυτό, αλλά στο ενδεχόμενο της χρήσης μιας συνάρτησης που θέλει τύπο δεδομένων ημερομηνίας και του δίνουμε αλφαριθμητική μεταβλητή δε θα εκτελεστεί σωστά.

Επιπλέον, εφόσον προσδιοριστούν οι τύποι των δεδομένων, η αποθήκευσή τους πραγματοποιείται από το πρόγραμμα σύμφωνα με τις κατηγορίες, δηλαδή τα πεδία και τους τύπους των δεδομένων. Οι σχέσεις μεταξύ των δομών-πινάκων όμως δημιουργούνται από το χρήστη. Εάν ο χρήστης επιθυμεί τα δεδομένα του να δομηθούν σύμφωνα με το σχήμα του αστερά ή της χιονονιφάδας είναι εφικτό, αρκεί

να μπορεί ο χρήστης να διακρίνει σωστά τις ιεραρχίες και τις σχέσεις μεταξύ των πινάκων των δεδομένων. Τις περισσότερες φορές το πρόγραμμα θα εμφανίσει μήνυμα σφάλματος στη δημιουργία μη κατάλληλης σχέσης μεταξύ των πινάκων, όπως την απουσία κάποιου πεδίου με μοναδικό χαρακτηριστικό (id ή κλειδί), αλλά στην εσφαλμένη δημιουργία σχέσης βάση λάθους της κρίσης του χρήστη δε θα αποδώσει στους μετέπειτα υπολογισμούς.

Πέρα από τις δυνατότητες των τροποποιήσεων των δεδομένων και τις DAX λειτουργίες που αναφέρθηκαν παραπάνω, μερικές από τις μορφές απεικονίσεων περιλαμβάνουν την επιλογή της χρήσης αναλυτικών στοιχείων. Με τη χρήση των αναλυτικών στοιχείων, μας δίνεται η δυνατότητα να προσθέσουμε δυναμικές αναφορικές γραμμές στα γραφήματά μας, οι οποίες δίνουν έμφαση σε σημαντικές τάσεις ή πληροφορίες. Για παράδειγμα, σε ένα ομαδοποιημένο διάγραμμα ράβδων (Clustered bar chart) (βλ. Εικόνα 3.8) ή σε ένα ομαδοποιημένο διάγραμμα στηλών (Clustered column chart) (βλ. Εικόνα 3.7) θα μπορούσαμε να προσθέσουμε μια δυναμική γραμμή αναφοράς που να έδειχνε το μέσο όρο των πωλήσεων ή κάποιου άλλου πεδίου. Έτσι, τα αποτελέσματα που μας παρουσιάζουν τα γραφήματα θα μπορούσαν να εξεταστούν με μία πιο σύνθετη οπτική. Επιπλέον, το πρόγραμμα διαθέτει τη δημιουργία δυναμικών γραμμών αναφοράς που δείχνουν το μέγιστο, το ελάχιστο, τη μέση τιμή και το μέσο όρο, όπου μπορούν να βοηθήσουν σημαντικά στην ερμηνεία-ανάλυση των γραφημάτων. Σκεφτείτε την περίπτωση ενός γραφήματος με τις πωλήσεις ανά κατηγορία των προϊόντων σε μία συγκεκριμένη περιοχή. Η ύπαρξη μερικών δυναμικών γραμμών αναφοράς θα μπορούσαν να ερμηνεύσουν την καλή ή την κακή πορεία μιας κατηγορίας προϊόντων. Την πληροφορία που προσδίδουν στις απεικονίσεις, αλλά και στο χρήστη θα τη δούμε πιο αναλυτικά στο τέταρτο κεφάλαιο.

Μία ακόμη δυνατότητα που διαθέτει το πρόγραμμα είναι η πρόβλεψη (Forecast), όπου με τη χρήση αυτού του εργαλείου ενσωματώνονται οι τάσεις και η εποχικότητα στα γραφήματα των απεικονίσεών μας. Πιο συγκεκριμένα, το εργαλείο της πρόβλεψης στο πρόγραμμα βασίζεται σε μια σειρά μεθόδων για την πρόβλεψη χρονοσειρών, όπου λέγεται εκθετική εξομάλυνση. Η εκθετική εξομάλυνση αποτελεί μια τεχνική πρόβλεψης, η οποία στηρίζεται στις πιο πρόσφατες παρατηρήσεις και θεωρεί ότι έχουν μεγαλύτερη σημασία, γι' αυτό το λόγο κατά την εφαρμογή της δίνει περισσότερη βαρύτητα σε αυτές. Το πλεονέκτημα αυτής της μεθόδου είναι ότι μειώνει αρκετά τους παράγοντες που μπορούν να παραμορφώσουν τα αποτελέσματα, όπως ο θόρυβος, και παράλληλα επιτυγχάνει στην αποτελεσματική καταγραφή των τάσεων.

Για να γίνει πιο κατανοητή η ανάλυση των χρονολογικών σειρών, ας δούμε πρώτα με ποια δεδομένα λειτουργεί και τι δεδομένα είναι αυτά. Τα δεδομένα χρονοσειρών αποτελούν δεδομένα που έχουν συλλεχθεί σε ομοιόμορφα τακτά χρονικά διαστήματα. Για παράδειγμα, η καταγραφή των πωλήσεων για ένα ή περισσότερα προκαθορισμένα διαστήματα ή η καταγραφή ευρύτερα μιας μεταβλητής, η οποία μετράται σταθερά σε συγκεκριμένους χρόνους με σκοπό την καταγραφή και την ανίχνευση των αλλαγών. Με την ανάλυση εξετάζονται αυτά τα δεδομένα και εξάγονται σημαντικά στατιστικά στοιχεία ή καινούρια δεδομένα. Η συλλογή τέτοιων δεδομένων στηρίζεται στη διαδοχική καταγραφή των τιμών και όχι τυχαία, θεωρώντας ότι αντιπροσωπεύουν μετρήσεις για ίσα χρονικά διαστήματα. Η ανάλυση των χρονοσειρών αποτελεί μια μέθοδο πρόβλεψης, η οποία στηρίζεται στο τι συνέβη στο παρελθόν για να προβλέψει τα γεγονότα του μέλλοντος.

Όσον αφορά την εποχικότητα, ορίζεται ως τα βήματα που απαιτούνται ώστε να ολοκληρωθεί ένας κύκλος δεδομένων, δηλαδή ένα φαινόμενο που εμφανίζεται επανειλημμένα σε όλες τις χρονολογικές σειρές. Για παράδειγμα, αν κατά την καταγραφή των πωλήσεων παρατηρείται μια αλλαγή που επαναλαμβάνεται κατά τον ίδιο τρόπο μέσα στον ίδιο κύκλο, δηλαδή στο ίδιο χρονικό διάστημα, τότε οι χρονολογικές σειρές περιλαμβάνουν εποχικότητα ή αλλιώς περιοδικότητα. Η εκτίμηση της εποχικότητας αποτελεί βασικό βήμα στην πρόβλεψη, διότι έχει ισχυρό αντίκτυπο στα αποτελέσματά της. Οι πιο γνωστοί αλγόριθμοι πρόβλεψης χρονολογικών σειρών είναι η εκθετική εξοικονόμηση χρόνου (ETS) και ο αυτορυθμιζόμενος ολοκληρωμένος κινητός μέσος όρος (ARIMA), οι οποίοι χρειάζονται την εποχικότητα ως είσοδο.

Για τον προσδιορισμό της εποχικότητας, το πρόγραμμα διαθέτει την επιλογή στον χρήστη να την εκτιμήσει από μόνος του, τοποθετώντας τον αριθμό των σημείων ανάλογα με το χρονικό διάστημα των δεδομένων και το χρονικό διάστημα της πρόβλεψης. Βέβαια, το πρόγραμμα εκτελεί κάποιες μεθόδους ώστε να ανιχνεύσει κατάλληλα την εποχικότητα και να προσαρμοστεί σωστά το μοντέλο. Με την εφαρμογή αυτών των μεθόδων κανονικοποιούνται οι τιμές και προσδιορίζονται οι συχνότητες και τα όρια των διαστημάτων των δεδομένων, σύμφωνα με τα χρονικά διαστήματα της καταγραφής, και προκύπτουν οι υποψήφιες περίοδοι. Στη συνέχεια, επιλέγεται η καταλληλότερη με γνώμονα τον συντελεστή αυτό-συσχέτισης του Pearson και Spearman. Στην περίπτωση όμως που ο χρήστης δώσει άλλο αριθμό για την εποχικότητα, ο οποίος περιλαμβάνει καλύτερα διαστήματα εμπιστοσύνης, δίνεται προτεραιότητα σε αυτήν την επιλογή.

Έτσι, με την ανάλυση των χρονολογικών σειρών μπορούν να προσδιοριστούν κάποια πρότυπα, τα οποία θα προεκταθούν στο μελλοντικό χρονικό διάστημα. Η χρήση δηλαδή αυτής της ανάλυσης είναι για όταν επιθυμούμε να προσδιορίσουμε την αιτία ενός φαινομένου που παρατηρήθηκε ή να προβλέψουμε τη "συμπεριφορά" αυτού του φαινομένου στο μέλλον. Για παράδειγμα, η ανάλυση αυτή μπορεί να μας βοηθήσει να δώσουμε απάντηση σε ερωτήματα όπως "Πώς εξελίσσονται οι πωλήσεις μας μέσω διαδικτύου;", "Μέχρι ποιο σημείο μπορούμε να βελτιώσουμε την παραγωγική διαδικασία για αυτό το προϊόν μας το επόμενο χρονικό διάστημα;", "Ποια θα είναι η ζήτηση μιας κατηγορίας των προϊόντων μας το επόμενο έτος;" Η μορφή απεικόνισης που χρησιμοποιείται για την ανάλυση τέτοιων δεδομένων είναι το γράφημα γραμμής, το οποίο είναι το καταλληλότερο για τα πεδία χρόνου.

Όσον αφορά την ανίχνευση των τάσεων, η ομαδοποίηση ή συσταδοποίηση μπορεί να βοηθήσει στην ανακάλυψη προτύπων στα δεδομένα μας. Στο Power BI Desktop υπάρχουν δύο τρόποι ώστε να πραγματοποιηθεί αυτό. Ο πρώτος είναι χρησιμοποιώντας διάγραμμα διασποράς (Scatter chart) όταν επιθυμούμε να εξετάσουμε δύο διαφορετικές μεταβλητές, τις οποίες χρησιμοποιούμε με σκοπό να βρούμε τις συστάδες ή τα σύνολα κι ο δεύτερος χρησιμοποιώντας πίνακα, ο οποίος περιλαμβάνει μεγάλο πλήθος μεταβλητών, όπου προσπαθούμε να ανακαλύψουμε την ύπαρξη "κρυμμένων" ομάδων εξετάζοντας τις μεταβλητές.

Για τη δημιουργία των συστάδων και τον προσδιορισμό των σημείων των δεδομένων σε αυτές, το πρόγραμμα διαθέτει δύο τρόπους συσταδοποίησης και είναι οι εξής:

• Ο αλγόριθμος K-means

Ο αλγόριθμος K-means, που αναφέρθηκε πιο αναλυτικά και στο δεύτερο κεφάλαιο, με τον οποίο ένα σημείο δεδομένων μπορεί να ανήκει σε μία μόνο συστάδα σύμφωνα

με τον υπολογισμό της απόστασής του από αυτήν. Ακόμη, ο συγκεκριμένος αλγόριθμος δουλεύει με συνεχείς μεταβλητές. Ωστόσο, στο πρόγραμμα έχει προσαρμοστεί να δουλεύει και με διακριτές τιμές χρησιμοποιώντας τις πιθανότητες. Μπορεί να δουλέψει με δύο τρόπους στα δεδομένα, είτε συνολικά είτε τμηματικά. Συνολικά σημαίνει να εξετάσει ολόκληρο το σύνολο των δεδομένων ώστε να τα ομαδοποιήσει και τμηματικά να ξεκινήσει εξετάζοντας ένα μέρος του συνόλου των δεδομένων κι εφόσον χρειαστεί θα επεκταθεί στα υπόλοιπα έως ότου καταλήξει στα βέλτιστα σύνολα.

• Η μέθοδος EM

Ο δεύτερος τρόπος συσταδοποίησης είναι η μέθοδος EM (Expectation Maximization), με την οποία ένα σημείο δεδομένων αντιστοιχεί πάντοτε σε πολλά σύνολα, δηλαδή συστάδες, και για κάθε συνδυασμό του σημείου με τις συστάδες μετράται μία πιθανότητα. Αποτελεί επίσης μια επαναληπτική διαδικασία με την οποία ο αλγόριθμος χρησιμοποιεί ένα παραδειγματικό μοντέλο συστάδας που ταιριάζει με τα δεδομένα και προσδιορίζει την πιθανότητά τους να υπάρχουν σε σημεία μέσα σε αυτήν. Η διαδικασία και οι υπολογισμοί αυτοί σταματούν να εκτελούνται όταν το αρχικό μοντέλο ταιριάζει με τον βέλτιστο τρόπο στα δεδομένα. Και σε αυτήν τη μέθοδο το πρόγραμμα διαθέτει την εφαρμογή της τμηματικά ή συνολικά, απλώς κατά τη συνολική εφαρμογή δεσμεύεται σημαντικά η μνήμη. Στην τμηματική η διαδικασία πραγματοποιείται γρηγορότερα, αλλά από θέμα απόδοσης είναι ισάξιες. Επιπλέον, αυτή η μέθοδος αποτελεί προεπιλογή στο πρόγραμμα, διότι είναι πιο προσεγγιστική η εφαρμογή της στα δεδομένα, δεν απαιτεί πολλές επαναλήψεις και προσφέρει περισσότερα σε σύγκριση με τον αλγόριθμο k-means.

Στο τέταρτο κεφάλαιο θα δούμε τη διαδικασία της συσταδοποίησης των δεδομένων που διαθέτει το πρόγραμμα για διάφορες περιπτώσεις στα δεδομένα μας.

Τέλος, η πληθώρα των εργαλείων και των δυνατοτήτων που διαθέτει το Power BI Desktop δεν περιορίζει καθόλου τον χρήστη, διότι το περιβάλλον του είναι ευχάριστο και του επιτρέπει την προσαρμογή και την επεξεργασία των δεδομένων του όπως επιθυμεί. Ακόμη, οι παρεμβάσεις του στον χρήστη εμφανίζοντάς του μηνύματα ενημέρωσης κατά την υλοποίηση των διαδικασιών του αποτελούν πολύτιμες για την αποτελεσματική του χρήση.

ΚΕΦΑΛΑΙΟ 4:

ΜΕΛΕΤΗ ΠΕΡΙΠΤΩΣΗΣ ΜΕ ΤΗ ΧΡΗΣΗ ΤΟΥ POWER BI

Στο παρόν κεφάλαιο θα πραγματοποιηθεί η ανάλυση κάποιων επιχειρηματικών δεδομένων, όπου η δομή τους και τα περιεχόμενά τους είναι διαμορφωμένα όπως είναι συνήθως στην πραγματικότητα. Τα δεδομένα μας αφορούν ένα κατάστημα λιανικής και συγκεκριμένα ένα σούπερ μάρκετ. Η εύρεσή τους πραγματοποιήθηκε μέσα από το διαδίκτυο. Στο πρώτο μέρος θα γίνει μια ευρύτερη περιγραφή του αρχείου των δεδομένων που έχουμε και στο δεύτερο μέρος θα πραγματοποιηθεί η φόρτωση των δεδομένων στο πρόγραμμα, η τροποποίηση-επεξεργασία τους και θα υλοποιηθεί η εξέταση-ανάλυσή τους με τη χρήση του Power BI Desktop¹.

4.1 ΠΕΡΙΓΡΑΦΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ

Από τα βασικότερα βήματα πριν προβούμε στην ανάλυση των δεδομένων είναι η κατανόηση των πληροφοριών και των σχέσεων που υπάρχουν μεταξύ τους. Ακόμη, η ερμηνεία τους παίζει σημαντικό ρόλο στην μελλοντική χρήση τους. Στη δική μας περίπτωση, τα δεδομένα που θα εξετάσουμε περιέχονται σε ένα αρχείο csv, το οποίο αφορά τις συναλλαγές που πραγματοποιήθηκαν. Κάθε γραμμή του αρχείου είναι μοναδική κι αφορά την αγορά συγκεκριμένου πελάτη και προϊόντος. Τα πεδία που περιλαμβάνει είναι τα εξής:

ItemCode: όπου είναι ο κωδικός του προϊόντος

sumAmount: το συνολικό χρηματικό ποσό της συναλλαγής

avgAmount: ο μέσος όρος του χρηματικού ποσού

avgStartAmount: ο μέσος όρος της αρχικής τιμής του προϊόντος

avgLineNr: η θέση (σειρά) που βρίσκεται το προϊόν στην απόδειξη

sumQty: η συνολική ποσότητα των προϊόντων

avgQty: ο μέσος όρος της συνολικής ποσότητας του προϊόντος

stdevQty: η τυπική απόκλιση της συνολικής ποσότητας του προϊόντος

NStores: το πλήθος των καταστημάτων όπου υπάρχει το προϊόν

NReceipts: το πλήθος των αποδείξεων για το συγκεκριμένο πελάτη για το συγκεκριμένο προϊόν

minDate: η ημερομηνία που ξεκίνησε η καταγραφή για το προϊόν

maxDate: η ημερομηνία που έληξε η καταγραφή για το προϊόν

¹ <https://powerbi.microsoft.com>

avgTime: ο μέσος όρος της ώρας που πραγματοποιήθηκε η συναλλαγή
 stdevTime: η τυπική απόκλιση της ώρας που πραγματοποιήθηκε η συναλλαγή
 avgWeekDay: η μέση μέρα που πραγματοποιήθηκε η συναλλαγή
 stdevWeekDay: η τυπική απόκλιση της μέρας που πραγματοποιήθηκε η συναλλαγή
 MG1Desc: η κατηγορία προϊόντος
 MG2Desc: η υποκατηγορία του MG1Desc
 MG3Desc: η υποκατηγορία του MG2Desc
 MG4Desc: η υποκατηγορία του MG3Desc
 ItemDescr: η περιγραφή του προϊόντος
 Recensy: πόσο πρόσφατα έγινε η συναλλαγή, δηλαδή πόσο πρόσφατα ήρθε ο συγκεκριμένος πελάτης για το συγκεκριμένο προϊόν
 Frequency: πόσο συχνά ήρθε ο συγκεκριμένος πελάτης για το συγκεκριμένο προϊόν στο χρονικό διάστημα καταγραφής
 Monetary: το συνολικό χρηματικό ποσό από το συγκεκριμένο πελάτη για το προϊόν
 pctAmount: ποσοστό χρημάτων από το συγκεκριμένο πελάτη για το συγκεκριμένο προϊόν

4.2 ΠΡΟ-ΕΠΕΞΕΡΓΑΣΙΑ ΤΩΝ ΔΕΔΟΜΕΝΩΝ

Όσον αφορά την προ-επεξεργασία των δεδομένων, κατά την φόρτωσή τους στο πρόγραμμα στην προεπισκόπηση που διαθέτει το πρόγραμμα παρατηρήσαμε πως υπήρχαν πολλά πεδία αριθμητικών τιμών με τελείες αντί για κόμμα, όπου το πρόγραμμα τους έδινε τον τύπο δεδομένων κειμένου. Ακόμη, για αυτό τον λόγο παρέλειπε και πολλές λανθασμένες τιμές που υπήρχαν στα δεδομένα, διότι υπήρχε μεγάλο πλήθος κατηγορηματικών μεταβλητών σε πεδία αριθμητικών κι αντίστροφα.

Χρησιμοποιώντας το Query Editor που διαθέτει το πρόγραμμα, διαμορφώσαμε τους τύπους των δεδομένων όπως πρέπει να είναι, δηλαδή για το SumAmount τον τύπο δεκαδικός αριθμός και για το MG1Desc τον τύπο κειμένου. Ακόμη, αφού πραγματοποιήσαμε τον έλεγχο όλων των τύπων των μεταβλητών, τα διαμορφώσαμε ώστε να έχουν συγκεκριμένα ψηφία για τους μετέπειτα υπολογισμούς τους. Όπως αναφέρθηκε και στο προηγούμενο κεφάλαιο ο προσδιορισμός των τύπων των δεδομένων αποτελεί βασικό παράγοντα για την ανάλυσή τους στη συνέχεια, διότι μπορεί το πρόγραμμα να τους επιλέγει αυτόματα με γνώμονα τις μεταβλητές που περιλαμβάνουν τα πεδία, αλλά μία λανθασμένη επιλογή μπορεί να οδηγήσει στην αδυναμία εφαρμογής των DAX λειτουργιών και των υπολογισμών ευρύτερα.

Όσον αφορά τις λανθασμένες τιμές ή τις κενές που περιελάμβανε το αρχείο μας, οι οποίες ήταν σημαντικά πολλές, κατά την εφαρμογή των σωστών τύπων των δεδομένων μετατρέπονταν σε σφάλματα, τα οποία θα παρέμεναν και μετά τη φόρτωσή τους. Αυτό σημαίνει πως τα αποτελέσματα που θα παίρναμε αργότερα θα είχαν μεγάλη απόκλιση από τα πραγματικά. Για να αποφύγουμε αυτό το γεγονός, εξετάσαμε τα σφάλματα για κάθε πεδίο και τα μετατρέψαμε σε null. Ακόμη, το πρόγραμμα μας διέθετε την αυτόματη διόρθωση λαθών, κατά την οποία αφαιρούνταν

ολόκληρη η γραμμή της καταγραφής. Κάτι τέτοιο όμως θα μας στερούσε μεγάλο μέρος πληροφοριών από άλλες στήλες που δεν περιελάμβαναν τόσα λάθη.

Μία άλλη τεχνική που διαθέτει το πρόγραμμα για τη διόρθωση των λαθών, είναι να δώσει προσεγγιστικές τιμές το ίδιο το πρόγραμμα στις μεταβλητές που είναι λανθασμένες. Αυτό θα μας χρησίμευε μέχρι ένα συγκεκριμένο ποσοστό λαθών, αλλά όπως προείπαμε το πλήθος των λαθών είναι μεγάλο όπου η εφαρμογή μιας τέτοιας τεχνικής θα είχε μεγάλη επιρροή στους μετέπειτα υπολογισμούς μας και κατά συνέπεια στα αποτελέσματά μας. Ακόμη, θα μπορούσαμε για ένα πλήθος τιμών να προσδώσουμε ένα βάρος, το οποίο θα προσδιόριζε στο πρόγραμμα σε ποιες τιμές να δώσει περισσότερο βαρύτητα και σε ποιες όχι. Και σε αυτήν την περίπτωση όμως θα συμπεριλαμβανόταν αυτές οι τιμές στους υπολογισμούς και θα ασκούσαν επιρροή στα αποτελέσματά μας. Οπότε κρατήσαμε τις προβληματικές τιμές που μετατρέψαμε με null, ώστε να μη χάσουμε τις πληροφορίες και από τα υπόλοιπα πεδία. Βέβαια, υπήρχαν πεδία όπου το ποσοστό των λαθών τους ήταν κοντά στο μισό του συνόλου των καταγραφών, για αυτόν τον λόγο δε χρησιμοποιήθηκαν. Ένας υπολογισμός σε ένα τέτοιο πεδίο δε θα μπορούσε να αντιπροσωπεύσει το σύνολο των δεδομένων παρά μόνο ένα δείγμα.

4.3 ΕΠΙΧΕΙΡΗΜΑΤΙΚΗ ΑΝΑΛΥΣΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΜΕ ΤΗ ΧΡΗΣΗ ΑΠΕΙΚΟΝΙΣΕΩΝ

Αφού περιγράψαμε το περιεχόμενο των δεδομένων του αρχείου μας και την προεπεξεργασία τους κατά την φόρτωσή τους στο πρόγραμμα, μπορούμε να ξεκινήσουμε τη διαδικασία ανάλυσής τους με τη χρήση των απεικονίσεων που μας διαθέτει το πρόγραμμα.

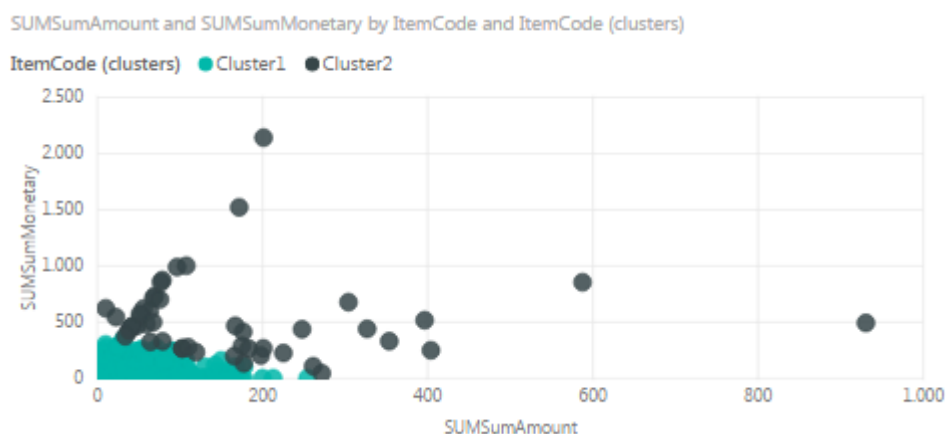
4.3.1 Συσταδοποίηση με τον κωδικό του προϊόντος

Θα ξεκινήσουμε με την τμηματοποίηση-συσταδοποίηση των δεδομένων μας σύμφωνα με το προϊόν και συγκεκριμένα με τον κωδικό του προϊόντος. Εφόσον όμως το αρχείο μας περιλαμβάνει μοναδικές εγγραφές (δηλαδή αφορά έναν πελάτη κι ένα προϊόν), σημαίνει ότι υπάρχουν πολλές συναλλαγές με τον ίδιο κωδικό προϊόντος και θα πρέπει να υπολογίσουμε τα συνολικά μεγέθη των πεδίων που το αφορούν, ώστε να βγάλουμε τις κατηγορίες-συστάδες που υπάρχουν στα δεδομένα μας.

Για να υπολογίσουμε τις συνολικές τιμές από το αρχείο μας, θα δημιουργήσουμε έναν πίνακα, ο οποίος με τη χρήση DAX συναρτήσεων, θα φιλτράρει τις καταγραφές των συναλλαγών σύμφωνα με τον κωδικό του προϊόντος και θα κρατάει αθροίζοντας τις συνολικές τιμές των πεδίων για κάθε προϊόν. Έτσι, θα έχουμε έναν πίνακα που θα περιλαμβάνει τον κωδικό του προϊόντος, το συνολικό ποσό και τη συνολική ποσότητα των συναλλαγών που το περιελάμβαναν. Ακόμη, θα κρατάει το συνολικό ποσό που βγάλαμε για κάθε προϊόν, την παλαιότερη και την πιο πρόσφατη ημερομηνία καταγραφής του. Ως κλειδί σύνδεσης του πίνακα που φτιάξαμε με το αρχείο των συναλλαγών θα έχουμε τον κωδικό του προϊόντος με τη σχέση ένα προς πολλά, διότι ένα προϊόν υπάρχει σε πολλές συναλλαγές. Έτσι, θα μπορούμε να έχουμε πρόσβαση και στα υπόλοιπα πεδία που αφορούν τις συναλλαγές. Σε αυτό το σημείο χρησιμοποιώντας αυτόν τον πίνακα μπορούμε να κατηγοριοποιήσουμε τα δεδομένα μας σύμφωνα με το προϊόν.

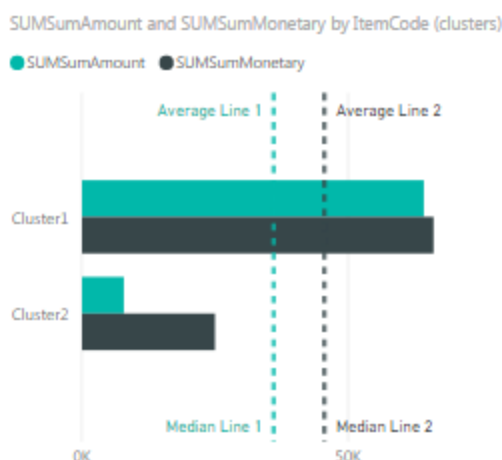
• Clustering by ItemCode

Επιλέγουμε για απεικόνιση το διάγραμμα διασποράς (Scatter Chart) και τοποθετούμε στον άξονα X το πεδίο SUMSumAmount και στον άξονα Y το πεδίο SUMSumMonetary. Ως πεδίο πληροφοριών για την κατηγοριοποίηση τοποθετούμε το ItemCode. Στη συνέχεια, μας διαθέτει το πρόγραμμα την επιλογή του πλήθους των συστάδων ή την αυτόματη, όπου το πρόγραμμα θα δημιουργήσει τις συστάδες σύμφωνα με τον αλγόριθμο που χρησιμοποιεί, τον οποίον αναφέραμε στο προηγούμενο κεφάλαιο. Επιλέγοντας την αυτόματη δημιουργία συστάδων, το πρόγραμμα δημιουργεί δύο συστάδες και παρουσιάζονται στην Εικόνα 4.1.



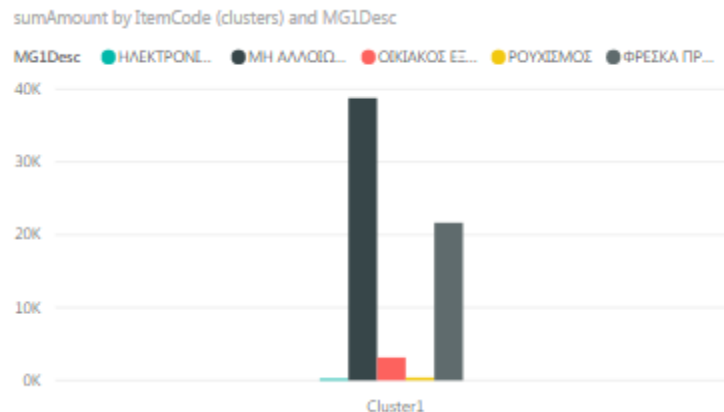
Εικόνα 4.1

Βλέπουμε ότι δημιούργησε δύο ομάδες προϊόντων σύμφωνα με το συνολικό χρηματικό ποσό που προέρχεται από την αγορά τους και το συνολικό ποσό κέρδους των συναλλαγών για τα προϊόντα. Η συστάδα 1 (cluster 1) είναι καλύτερα ομαδοποιημένη από ότι η δεύτερη, η οποία περιλαμβάνει αρκετά σημεία σε απόσταση (outliers), οπότε θα εξετάσουμε τις συστάδες με ένα ομαδοποιημένο διάγραμμα ράβδων και με γραμμές αναλυτικών στοιχείων για να δούμε αν όντως πρέπει να εξετάσουμε τη συστάδα 1. Όπως βλέπουμε στην Εικόνα 4.2, οι συνολικές τιμές της πρώτης συστάδας κυμαίνονται πάνω από τους μέσους όρους και τις μέσες τιμές, ενώ της δεύτερης όχι. Άρα, τα προϊόντα αυτής της συστάδας αξίζουν περαιτέρω διερεύνηση.



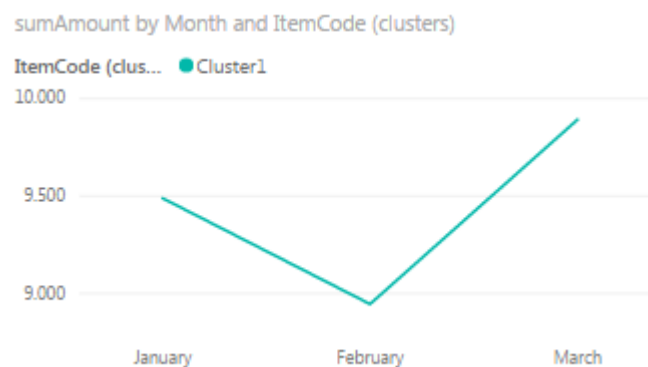
Εικόνα 4.2

Για να πάρουμε περισσότερες πληροφορίες για αυτή την ομάδα που δημιουργήθηκε, θα επιλέξουμε το πεδίο της πρώτης συστάδας και το πεδίο MG1Desc από το αρχείο μας, όπου μας δίνει το ευρύτερο επίπεδο της κατηγορίας των προϊόντων και χρησιμοποιώντας το ομαδοποιημένο διάγραμμα στηλών (Clustered Column Chart) θα δούμε ποια κατηγορία προϊόντων έχει τις υψηλότερες πωλήσεις σε αυτή την ομάδα. Όπως βλέπουμε στην Εικόνα 4.3, η κατηγορία ΜΗ ΑΛΛΟΙΩΣΙΜΑ των προϊόντων είναι αυτή με τις υψηλότερες πωλήσεις και επόμενη είναι η κατηγορία με τα ΦΡΕΣΚΑ ΠΡΟΙΟΝΤΑ. Οι υπόλοιπες τρεις κατηγορίες σημειώνουν χαμηλά επίπεδα πωλήσεων, οπότε θα ερευνήσουμε περισσότερο τις δύο πρώτες κατηγορίες.



Εικόνα 4.3

Ακόμη, θα μπορούσαμε να δούμε την πορεία της συγκεκριμένης κατηγορίας των προϊόντων της συστάδας σύμφωνα με το αρχείο συναλλαγών που τα περιλαμβάνει και το χρονικό διάστημα των καταγραφών τους. Έτσι, με τη χρήση ενός γραφήματος γραμμής, όπως φαίνεται στην Εικόνα 4.4, επιλέγουμε τα πεδία sumAmount, MinDate, MaxDate και το πεδίο της συστάδας 1. Σε αυτό το σημείο είναι απαραίτητο να δημιουργούμε ένα Boolean πεδίο, όπου επιστρέφει την τιμή true όταν η ημερομηνία έναρξης και λήξης της καταγραφής είναι ίδια, δηλαδή η συναλλαγή είναι ημερήσια, και το χρησιμοποιούμε ως φίλτρο στην απεικόνιση για να δούμε την κατάλληλη πορεία των ημερήσιων πωλήσεων των προϊόντων. Επιπλέον, επιλέγουμε ως φίλτρο και το πεδίο MG1Desc, ώστε να περιλαμβάνονται μόνο τα ΜΗ ΑΛΛΟΙΩΣΙΜΑ προϊόντα.



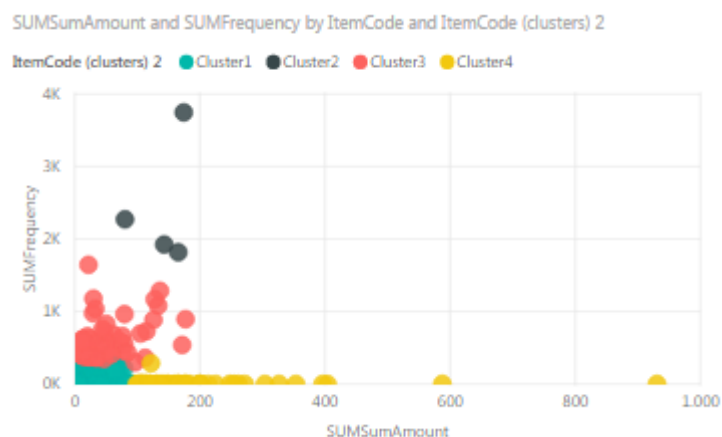
Εικόνα 4.4

Παρατηρώντας την Εικόνα 4.4, βλέπουμε πως σύμφωνα με τις συναλλαγές που πραγματοποιήθηκαν οι πωλήσεις αυτών των προϊόντων είχαν πτώση τον Φεβρουάριο, αλλά σημείωσαν σημαντική άνοδο έως το Μάρτιο.

Σε αυτό το σημείο αξίζει να εξετάσουμε τα προϊόντα όσον αφορά την συχνότητα που αγοράστηκαν. Στη συνέχεια θα χρησιμοποιήσουμε ως φίλτρο την παραπάνω συστάδα, ώστε να δούμε ποια προϊόντα από αυτά που την απαρτίζουν αγοράζονται συχνότερα, δηλαδή έχουν την υψηλότερη ζήτηση.

Για να χρησιμοποιήσουμε το πεδίο της συχνότητας από το αρχείο των συναλλαγών, θα πρέπει να το προσθέσουμε στον πίνακα των προϊόντων που δημιουργήσαμε παραπάνω. Με τη χρήση των DAX συναρτήσεων θα πάρουμε το άθροισμα όλων των συχνοτήτων από τον πίνακα των συναλλαγών για κάθε κωδικό προϊόντος. Έτσι, στον πίνακα των προϊόντων θα έχουμε τη στήλη με τη συχνότητα, όπου το ύψος της τιμής της θα μας δείχνει πόσο συχνά αγοράστηκε το προϊόν. Όσο μεγαλύτερη θα είναι αυτή η τιμή για ένα προϊόν, τόσο συχνότερα αγοράστηκε.

Αφού δημιουργήσαμε το πεδίο της συχνότητας, θα δημιουργήσουμε ένα διάγραμμα διασποράς, ώστε να δούμε τη σχέση της συνολικής συχνότητας των πωλήσεων με το συνολικό ποσό των πωλήσεων για τα προϊόντα. Όπως βλέπουμε στην Εικόνα 4.5 δεν υπάρχει κάποια σχέση μεταξύ τους, αλλά το πρόγραμμα με την αυτόματη επιλογή μας δημιούργησε τέσσερις συστάδες. Η συστάδα δύο περιλαμβάνει τα προϊόντα που αγοράζονται συχνότερα, αλλά δεν έχουν τα υψηλότερα επίπεδα στις πωλήσεις μας σε σύγκριση με τα υπόλοιπα. Αυτό το αποτέλεσμα δεν αποτελεί μη ερμηνεύσιμο, διότι δικαιολογείται ανάλογα με το είδος και την τιμή του προϊόντος. Έτσι, θα πάρουμε τη συστάδα της Εικόνας 4.3 και με τη χρήση ενός πίνακα απεικόνισης θα πάρουμε την ομάδα των προϊόντων που έχουν τις υψηλότερες πωλήσεις.



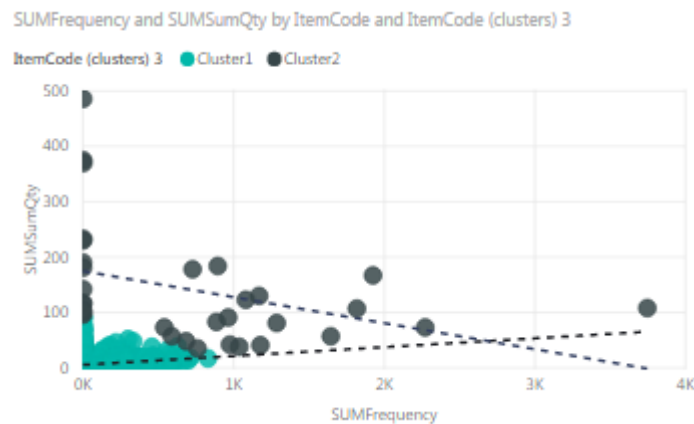
Εικόνα 4.5

Για να δούμε στον πίνακα ποιο προϊόν αγοράζεται συχνότερα από αυτή την ομάδα, θα χρησιμοποιήσουμε δύο φίλτρα. Το ένα θα είναι το πεδίο με τις κατηγορίες των προϊόντων, όπου και θα κρατάει τις κατηγορίες ΜΗ ΑΛΛΟΙΩΣΙΜΑ και ΦΡΕΣΚΑ ΠΡΟΙΟΝΤΑ. Το άλλο φίλτρο θα είναι το πεδίο των συστάδων που δημιουργήσαμε στην Εικόνα 4.5, όπου και θα επιλέξουμε την εμφάνιση των αποτελεσμάτων για τα προϊόντα που αγοράζονται συχνότερα, δηλαδή τη συστάδα 2 της Εικόνας 4.5. Τα αποτελέσματα που μας δίνει ο πίνακας (βλ. Εικόνα 4.6), δείχνουν ποια προϊόντα και τη συχνότητα που αγοράζονται από την συστάδα 1 της Εικόνας 4.3. Δηλαδή, από την ομάδα των προϊόντων της συστάδας 1, που σημείωσε τις περισσότερες πωλήσεις στα ΜΗ ΑΛΛΟΙΩΣΙΜΑ και στα ΦΡΕΣΚΑ ΠΡΟΙΟΝΤΑ, τα προϊόντα ΤΟΣΤ ΜΙΝΙ ΚΑΡΑΜΟΛΕΓΚ ΣΤΑΡ ST-Ο και ΓΑΛΑ ΥΨΗΛ ΜR GRAND 3 είχαν τη μεγαλύτερη συχνότητα, δηλαδή αγοράστηκαν περισσότερο.

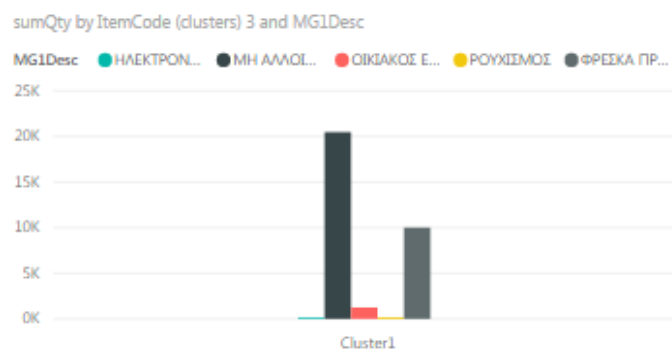
ItemCode	MG1Desc	ItemDescr	SUMFrequency	sumAmount
2494029	ΜΗ ΑΛΛΟΙΩΣΙΜΑ	ΤΟΣΤ ΜΙΝΙ ΚΑΡΑΜΟΛΕΓΚ ΣΤΑΡ ST-0	2.272,00	80,36
3066248	ΦΡΕΣΚΑ ΠΡΟΙΟΝΤΑ	ΓΑΛΑ ΥΨΗΛ ΜR GRAND 3	1.925,00	143,00
Total				223,36

Εικόνα 4.6

Στη συνέχεια όμως, ας εξετάσουμε αν σχετίζεται και η ποσότητα των προϊόντων με τη συχνότητα που αγοράζονται. Θα δημιουργήσουμε ένα διάγραμμα διασποράς, με τα πεδία SUMSumQty, SUMFrequency και ItemCode. Το πρόγραμμα μας δίνει δύο συστάδες (βλ. Εικόνα 4.7), όπου η πρώτη δείχνει ότι υπάρχει κάποια σχέση μεταξύ των προϊόντων, ενώ η δεύτερη ομάδα δείχνει ότι έχουν αρνητική σχέση. Για να εξετάσουμε περισσότερο την συστάδα 1, θα χρησιμοποιήσουμε ένα ομαδοποιημένο διάγραμμα στηλών (Clustered Column Chart), όπου θα προσθέσουμε πέρα από το πεδίο των συστάδων που δημιουργήθηκαν και το πεδίο με τις κατηγορίες των προϊόντων από το αρχείο συναλλαγών. Τα αποτελέσματα του διαγράμματος αυτού (βλ. Εικόνα 4.8) μας δείχνουν πως τα προϊόντα, που η ποσότητά τους έχει κάποια σχέση με τη συχνότητα της αγοράς τους, ανήκουν στις κατηγορίες ΜΗ ΑΛΛΟΙΩΣΙΜΑ και ΦΡΕΣΚΑ ΠΡΟΙΟΝΤΑ



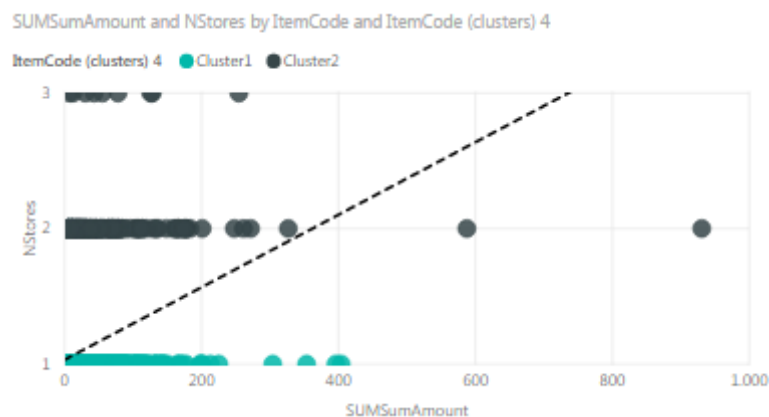
Εικόνα 4.7



Εικόνα 4.8

Με τη βοήθεια ενός απλού πίνακα απεικόνισης είδαμε πως το πλήθος τους είναι μεγάλο για να εξάγουμε συμπεράσματα. Συγκεκριμένα τα ΜΗ ΑΛΛΟΙΩΣΙΜΑ είναι 13.425 και τα ΦΡΕΣΚΑ ΠΡΟΙΟΝΤΑ είναι 6.423. Οπότε θα κρατήσουμε τον πίνακα για να τον φιλτράρουμε στην πορεία με κάποια άλλη ομάδα για τα προϊόντα ή για τον πελάτη.

Η επόμενη εξέτασή μας θα αφορά το συνολικό ποσό των πωλήσεων για κάθε προϊόν σε σχέση με τον αριθμό του καταστήματος που πωλείται. Με τη χρήση των DAX συναρτήσεων θα πάρουμε το αριθμό των καταστημάτων από τον πίνακα των συναλλαγών για κάθε κωδικό του προϊόντος. Στη συνέχεια, θα δημιουργήσουμε ένα διάγραμμα διασποράς, όπου ως πεδία θα τοποθετήσουμε το SUMSumAmount, το NStores που δημιουργήσαμε και τον ItemCode. Από το διάγραμμα (βλ. Εικόνα 4.9) βλέπουμε ότι δε μπορούμε να αντλήσουμε πληροφορία. Αφού το πρόγραμμα με την αυτόματη επιλογή έδωσε δύο συστάδες προσθέσαμε ένα αναλυτικό στοιχείο και συγκεκριμένα τη γραμμή τάσης για κάθε κατάσταση για να ερμηνεύσουμε τη διαφορετικότητα των συστάδων. Πράγματι, η συστάδα 1, όπου περιλαμβάνει τα προϊόντα που υπάρχουν στο ένα κατάστημα, δείχνει ότι το συνολικό ποσό των πωλήσεων των προϊόντων σχετίζεται με το κατάστημα όπου πωλούνται.



Εικόνα 4.9

Χρησιμοποιώντας την απεικόνιση ενός απλού πίνακα (βλ Εικόνα 4.10), επιλέξαμε το πεδίο των ομάδων που δημιουργήθηκαν, το συνολικό ποσό των πωλήσεων, τα καταστήματα και τη συχνότητα. Παρατηρούμε πως η συστάδα 1, όπου αφορά τα προϊόντα που πωλούνται στο ένα κατάστημα, έχει τις υψηλότερες πωλήσεις και το μεγαλύτερο άθροισμα της συχνότητας. Αυτό σημαίνει πως τα προϊόντα αυτά του καταστήματος 1 πωλούνται συχνότερα σε σύγκριση με των άλλων ομάδων.

ItemCode (clusters) 4	SUMSumAmount	NStores	SUMFrequency
Cluster1	58.660,11	1	91.747,50
Cluster2	12.696,65	2	17.283,67
Cluster2	735,03	3	2.056,30

Εικόνα 4.10

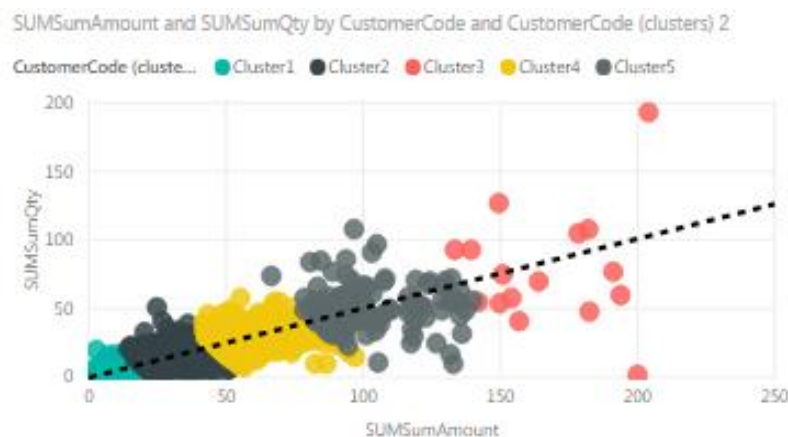
Αφού δημιουργήσαμε κάποιες ομάδες από την σκοπιά του προϊόντος κι εξάγαμε ευρύτερα κάποια συμπεράσματα για τα δεδομένα μας, θα συνεχίσουμε την ανάλυση δημιουργώντας ομαδοποιήσεις και από την πλευρά του καταναλωτή. Σε αυτό το σημείο είναι σημαντικό να αναφέρουμε πως οι ομάδες που δημιουργήθηκαν θα χρησιμοποιηθούν στη δημιουργία αναφορών στη συνέχεια

4.3.2 Συσταδοποίηση με τον κωδικό του πελάτη

• Clustering by CustomerCode

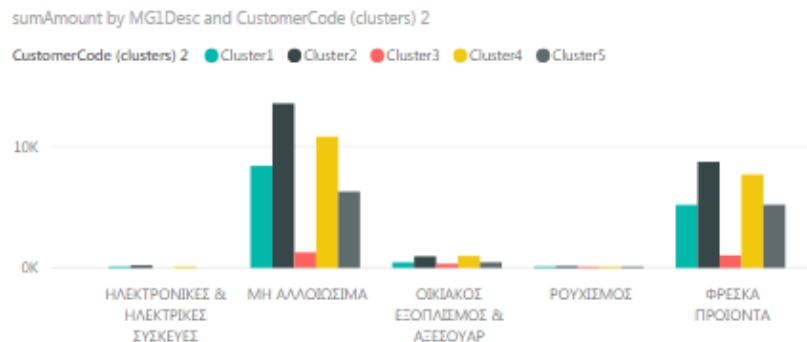
Πριν την ομαδοποίηση των δεδομένων μας σύμφωνα με τους πελάτες μας θα πρέπει να δημιουργήσουμε έναν συνολικό πίνακα που να αφορά αποκλειστικά τον πελάτη και να περιλαμβάνει τις συνολικές τιμές των μεταβλητών ανά μονάδα του. Έτσι, θα δημιουργήσουμε άλλον έναν πίνακα, ο οποίος με τη χρήση DAX συναρτήσεων, θα φιλτράρει τις καταγραφές των συναλλαγών σύμφωνα με τον κωδικό του πελάτη και θα κρατάει αθροίζοντας τις συνολικές τιμές των πεδίων για κάθε πελάτη. Τα περιεχόμενα αυτού του πίνακα θα είναι ο κωδικός του πελάτη, το συνολικό ποσό που έδωσε και η συνολική ποσότητα που αγόρασε σύμφωνα με το αρχείο των συναλλαγών. Ακόμη, θα περιλαμβάνει το συνολικό ποσό που βγάλαμε από κάθε πελάτη, την παλαιότερη και την πιο πρόσφατη ημερομηνία που πραγματοποίησε κάποια αγορά του. Ως κλειδί σύνδεσης του πίνακα που φτιάξαμε με το αρχείο των συναλλαγών θα έχουμε τον κωδικό του πελάτη με τη σχέση ένα προς πολλά, διότι ένας πελάτης μπορεί να πραγματοποιήσει πολλές συναλλαγές. Έτσι, όπως και με τον πίνακα των προϊόντων, θα μπορούμε να έχουμε πρόσβαση και στα υπόλοιπα πεδία. Σε αυτό το σημείο χρησιμοποιώντας αυτόν τον πίνακα μπορούμε να κατηγοριοποιήσουμε τα δεδομένα μας σύμφωνα με τον πελάτη.

Για την ομαδοποίηση, επιλέγουμε για απεικόνιση το διάγραμμα διασποράς (Scatter Chart) και τοποθετούμε στον άξονα X το πεδίο SUMSumAmount και στον άξονα Y το πεδίο SUMSumQty. Ως πεδίο πληροφοριών για την κατηγοριοποίηση τοποθετούμε το CustomerCode. Επιλέγοντας την αυτόματη δημιουργία συστάδων, το πρόγραμμα δημιουργεί πέντε συστάδες και παρουσιάζονται στην Εικόνα 4.11. Όπως βλέπουμε υπάρχει θετική συσχέτιση της ποσότητας των προϊόντων με τις αγορές των πελατών. Θα εξετάσουμε περαιτέρω τους καταναλωτές που περιλαμβάνονται στις συστάδες 1, 2 και 4, διότι είναι αυτές που δεν έχουν τιμές μακριά απ' το κέντρο τους.



Εικόνα 4.11

Όσον αφορά τις τρεις αυτές συστάδες, μας ενδιαφέρει αρχικά να δούμε τις κατηγορίες των προϊόντων που αγοράζουν. Για να το δούμε αυτό θα χρησιμοποιήσουμε την απεικόνιση ενός ομαδοποιημένου διαγράμματος στηλών. Στο συγκεκριμένο διάγραμμα (βλ. Εικόνα 4.12), παρατηρούμε ότι οι τρεις αυτές συστάδες πραγματοποιούν συναλλαγές, οι οποίες περιλαμβάνουν προϊόντα των κατηγοριών ΜΗ ΑΛΛΟΙΩΣΙΜΑ και ΦΡΕΣΚΑ ΠΡΟΙΟΝΤΑ.



Εικόνα 4.12

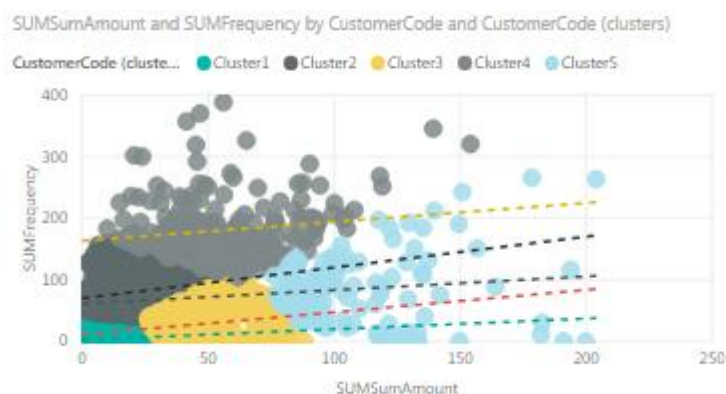
Για να εξάγουμε περισσότερα συμπεράσματα για αυτές τις ομάδες καταναλωτών, θα δημιουργήσουμε έναν matrix πίνακα (βλ. Εικόνα 13), όπου θα προσθέσουμε ως πεδία, το πεδίο των συστάδων που δημιουργήθηκε, τις κατηγορίες προϊόντων και τα συνολικά ποσά των πωλήσεων σύμφωνα με τις συναλλαγές που πραγματοποιήθηκαν. Παρατηρούμε πως η ομάδα των καταναλωτών της συστάδας 4 σημείωσε τις υψηλότερες πωλήσεις και στις δύο κατηγορίες, ενώ η συστάδα 5, όπου είχε αποστάσεις μεταξύ των σημείων της έρχεται δεύτερη στα ύψη των πωλήσεων. Αυτό οφείλεται στο ότι το πλήθος των συναλλαγών μπορεί να ήταν μικρό, αλλά τα ποσά των συναλλαγών υψηλά. Οπότε θα χρησιμοποιήσουμε τη συστάδα 4 ως φίλτρο στη συνέχεια.

CustomerCode (clusters) 2	ΜΗ ΑΛΛΟΙΩΣΙΜΑ	ΦΡΕΣΚΑ ΠΡΟΪΟΝΤΑ	Total
Cluster1	8.906,84	5.614,08	14.520,92
Cluster2	1.262,16	1.001,77	2.263,93
Cluster3	5.597,91	4.547,39	10.145,30
Cluster4	14.047,47	9.040,06	23.087,53
Cluster5	10.553,87	7.677,45	18.231,32
Total	40.368,25	27.880,75	68.249,00

Εικόνα 4.13

Σε αυτό το σημείο ας ομαδοποιήσουμε τους πελάτες σύμφωνα με τη συχνότητα και το σύνολο των αγορών που πραγματοποίησαν. Δημιουργώντας ένα διάγραμμα διασποράς με τον κωδικό του πελάτη, το συνολικό ποσό των πωλήσεων και το σκορ της συχνότητας που του αναλογεί. Όπως στον πίνακα με τα προϊόντα, έτσι και στον πίνακα που αφορά τον πελάτη πήραμε το άθροισμα των συχνοτήτων από το αρχείο των συναλλαγών. Το αποτέλεσμα του διαγράμματος με την αυτόματη δημιουργία συστάδων μας δίνει δύο μεγάλες συστάδες, όπου αρκετές παρατηρήσεις τους έχουν αποστάσεις μεταξύ τους. Οπότε θα επιλέξουμε εμείς τον αριθμό των συστάδων, ώστε να έχουμε πιο μικρά δείγματα για καλύτερη περεταίρω ανάλυση. Μετά από διάφορες δοκιμές, θα δώσουμε τον αριθμό πέντε για τη δημιουργία πέντε συστάδων, διότι αναλογικά όπως φαίνεται στο γράφημα (βλ. Εικόνα 4.14) πραγματοποιείται καλύτερη ομαδοποίηση.

Για την καλύτερη κατανόηση του γραφήματος και την ανίχνευση προτύπων προσθέσαμε αναλυτικές γραμμές, οι οποίες αναπαριστούν τις τάσεις των πελατών σύμφωνα με τη συχνότητα και το ποσό που πληρώνουν όταν πραγματοποιούν τις αγορές τους. Έτσι, βλέπουμε ότι η συστάδα 2 (μαύρη γραμμή) και η συστάδα 3 (κίτρινη γραμμή) σημειώνουν μια σημαντική σχέση. Είναι κι οι καλύτερα ομαδοποιημένες σύμφωνα με την εικόνα του γραφήματος.



Εικόνα 4.14

Για την καλύτερη ερμηνεία αυτών των αποτελεσμάτων θα δημιουργήσουμε έναν πίνακα που θα περιλαμβάνει τις ομάδες των καταναλωτών που δημιουργήθηκαν, τα συνολικά ποσά που έδωσαν και τη συνολική συχνότητά των αγορών τους. Όπως βλέπουμε στην Εικόνα 4.15 η συστάδα 2 έχει τη μεγαλύτερη συχνότητα, που σημαίνει ότι οι καταναλωτές που την απαρτίζουν πραγματοποιούν συναλλαγές συχνότερα από τους άλλους. Ακόμη, η συστάδα 3 αν και δεν έχει υψηλή συχνότητα σημειώνει σημαντικό ύψος στο συνολικό ποσό των συναλλαγών, δηλαδή αποτελείται από πελάτες που τα ποσά των συναλλαγών τους είναι υψηλότερα σε σύγκριση με τους άλλους, αλλά η συχνότητα των αγορών τους είναι χαμηλή.

CustomerCode (clusters)	SUMSumAmount	SUMFrequency
Cluster1	17.296,49	7.811,99
Cluster2	10.768,95	45.321,13
Cluster3	19.717,55	11.781,50
Cluster4	11.279,74	35.996,46
Cluster5	13.029,06	10.176,39

Εικόνα 4.15

Για να εξετάσουμε βαθύτερα αυτές τις δύο ομάδες θα χρησιμοποιήσουμε την απεικόνιση ενός matrix πίνακα με τα πεδία της συστάδας που δημιουργήθηκε, τις κατηγορίες των προϊόντων και τα συνολικά ποσά των συναλλαγών. Επειδή οι κατηγορίες των προϊόντων περιέχουν ιεραρχίες, θα εφαρμόσουμε την τεχνική της αναλυτικής καθόδου (Drill-Down) (βλ. Ενότητα 2.3.4), όπου μας τη διαθέτει αυτή η απεικόνιση, και θα μειώσουμε τα επίπεδα των ιεραρχιών ώστε να δούμε ποιες κατηγορίες προϊόντων αγοράζουν οι συχνότεροι και μη καταναλωτές μας. Όπως βλέπουμε στον πίνακα (βλ. Εικόνα 4.16), οι συχνότεροι καταναλωτές (συστάδα 2) και οι λιγότερο συχνοί (συστάδα 3) πραγματοποιούν περισσότερες αγορές στα ΜΗ ΑΛΛΟΙΩΣΙΜΑ και στα ΦΡΕΣΚΑ ΠΡΟΙΟΝΤΑ. Οι λιγότεροι συχνοί όπως είδαμε και παραπάνω έχουν τα μεγαλύτερα συνολικά ποσά σύμφωνα με τις συναλλαγές τους και συγκεκριμένα το μεγαλύτερο μέρος τους αφορά τα ΜΗ ΑΛΛΟΙΩΣΙΜΑ.

CustomerCode (clusters)	sumAmount
Cluster2	10.768,95
ΗΛΕΚΤΡΟΝΙΚΕΣ & ΗΛΕΚΤΡΙΚΕΣ ΣΥΣΚΕΥΕΣ	13,74
ΜΗ ΑΛΛΟΙΩΣΙΜΑ	6.660,43
ΟΙΚΙΑΚΟΣ ΕΞΟΠΛΙΣΜΟΣ & ΑΞΕΣΟΥΑΡ	344,22
ΡΟΥΧΙΣΜΟΣ	70,47
ΦΡΕΣΚΑ ΠΡΟΪΟΝΤΑ	3.680,09
Cluster3	19.717,55
ΗΛΕΚΤΡΟΝΙΚΕΣ & ΗΛΕΚΤΡΙΚΕΣ ΣΥΣΚΕΥΕΣ	129,82
ΜΗ ΑΛΛΟΙΩΣΙΜΑ	10.862,57
ΟΙΚΙΑΚΟΣ ΕΞΟΠΛΙΣΜΟΣ & ΑΞΕΣΟΥΑΡ	1.077,13
ΡΟΥΧΙΣΜΟΣ	112,93
ΦΡΕΣΚΑ ΠΡΟΪΟΝΤΑ	7.535,10
Total	30.486,50

Εικόνα 4.16

Για να εξάγουμε περισσότερες πληροφορίες για τα προϊόντα που αγοράζουν οι λιγότεροι συχνοί καταναλωτές (συστάδα 3), θα αφαιρέσουμε από την απεικόνιση του πίνακα τη συστάδα 2 χρησιμοποιώντας ως φίλτρο το πεδίο των συστάδων κι επιλέγοντας μόνο την εμφάνιση της τρίτης ομάδας. Επιπλέον, θα εφαρμόσουμε την τεχνική της συναθροιστικής ανόδου (Roll-up) (βλ. Ενότητα 2.3.4), ώστε να αυξήσουμε κατά ένα επίπεδο την ιεραρχία των κατηγοριών των προϊόντων. Τέλος, θα εφαρμόσουμε φθίνουσα ταξινόμηση στο πεδίο των συνολικών ποσών των αγορών κι έτσι θα δούμε ποια κατηγορία δεύτερου επιπέδου έχει τις περισσότερες συνολικές πωλήσεις σύμφωνα με τις συναλλαγές των λιγότερων συχνών πελατών. Τα αποτελέσματα που μας δίνει ο πίνακας (βλ. Εικόνα 4.17), δείχνουν πως οι λιγότερο συχνοί πελάτες πραγματοποιούνε περισσότερες αγορές σε προϊόντα παντοπωλείου.

CustomerCode (clusters)	sumAmount
Cluster3	19.717,55
ΠΑΝΤΟΠΩΛΕΙΟ	3.299,75

Εικόνα 4.17

Για να δούμε την υποκατηγορία του παντοπωλείου, όπου οι λιγότερο συχνοί καταναλωτές πραγματοποιούν τις περισσότερες αγορές, θα αυξήσουμε την ιεραρχία κατά δύο επίπεδα. Μετά από αυτήν την τροποποίηση τα αποτελέσματα του πίνακα (βλ. Εικόνα 4.18), δείχνουν πως τα προϊόντα της υποκατηγορίας ΠΡΩΙΝΑ και συγκεκριμένα τα Corn flakes-μούσλι αγοράζονται περισσότερο από την ομάδα των λιγότερο συχνών πελατών. Εάν κάνουμε την ίδια διαδικασία και για τους συχνούς πελάτες (συστάδα 2) βλέπουμε ότι κι εκείνοι αγοράζουν προϊόντα της υποκατηγορίας ΠΡΩΙΝΑ, απλώς σημειώνουν χαμηλότερες πωλήσεις.

CustomerCode (clusters)	sumAmount
Cluster3	3.299,75
ΜΗ ΑΛΛΟΙΩΣΙΜΑ	3.299,75
ΠΑΝΤΟΠΩΛΕΙΟ	3.299,75
ΠΡΩΙΝΑ	436,29
Corn flakes-μούσλι	346,87

Εικόνα 4.18

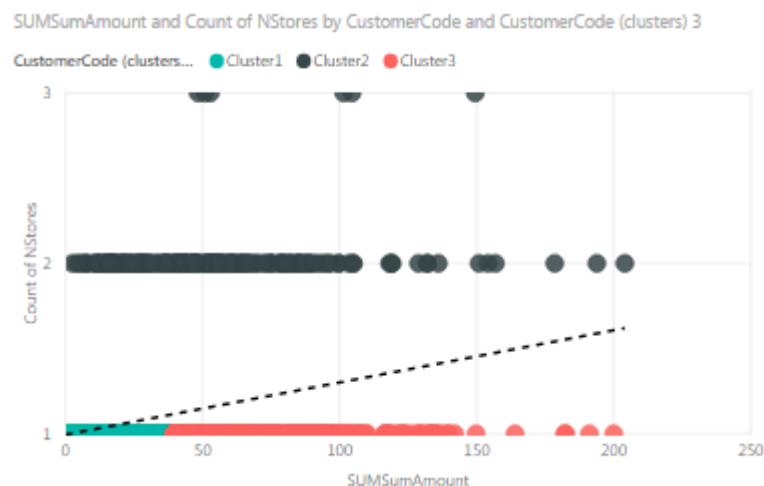
Για να εξακριβώσουμε τη διαφορά αυτών των δύο ομάδων, πέρα από το σύνολο των πωλήσεων θα χρησιμοποιήσουμε ως φίλτρο τη συστάδα 4 της Εικόνας 4.13 , όπου περιλαμβάνει την ομάδα των πελατών που σχετίζονται τα συνολικά ποσά των αγορών τους με τις ποσότητες που αγόρασαν και σημείωσαν τις περισσότερες πωλήσεις σύμφωνα με τις συναλλαγές. Τα αποτελέσματα του πίνακα (βλ. Εικόνα 4.19),

αλλάζουν μετά την εφαρμογή του φίλτρου και δείχνουν πως οι συχνοί και οι λιγότεροι συχνοί καταναλωτές πραγματοποίησαν περισσότερες πωλήσεις στην υποκατηγορία ΚΑΦΕΔΕΣ και συγκεκριμένα στους καφέδες καπουτσίνο-εσπρέσο με τη διαφορά ότι οι συχνοί καταναλωτές σημείωσαν λιγότερες αγορές από τους λιγότερο συχνούς. Ακόμη, βλέπουμε ότι οι λιγότεροι συχνοί πελάτες αγοράζουν περισσότερο καφέ στιγμής, ενώ οι συχνοί ελληνικό καφέ.

MG1Desc	Cluster2	Cluster3	Total
ΜΗ ΑΛΛΟΙΩΣΙΜΑ	1.097,97	1.521,37	2.619,34
ΠΑΝΤΟΠΩΛΕΙΟ	1.097,97	1.521,37	2.619,34
ΚΑΦΕΔΕΣ	163,72	266,93	430,65
Καφές καπουτσίνο-εσπρέσο	58,13	115,86	173,99
Καφές στιγμής	37,94	75,65	113,59
Καφές ελληνικός	43,85	54,69	98,54
Καφές φίλτρου	23,14	17,85	40,99
Φίλτρα καφέ	0,66	2,88	3,54

Εικόνα 4.19

Ας δούμε όμως με ποιον τρόπο ομαδοποιούνται οι πελάτες σε συνάρτηση με τις αγορές που πραγματοποιούν και τα καταστήματα που επιλέγουν. Από το διάγραμμα (βλ. Εικόνα 4.20) δε μπορούμε να αντλήσουμε πολλή πληροφορία. Αφού το πρόγραμμα με την αυτόματη επιλογή έδωσε τρεις συστάδες, προσθέσαμε ένα αναλυτικό στοιχείο και συγκεκριμένα τη γραμμή τάσης για κάθε κατάσταση για να ερμηνεύσουμε τη διαφορετικότητα των συστάδων. Βλέπουμε ότι υπάρχει μία τάση στο κατάστημα ένα, όπου περιλαμβάνονται δύο ομάδες καταναλωτών.



Εικόνα 4.20

Για να ερμηνεύσουμε αυτή την τάση θα δημιουργήσουμε έναν matrix πίνακα με τα πεδία της ομαδοποίησης που δημιουργήθηκαν, τις κατηγορίες των προϊόντων, τα καταστήματα και τα συνολικά ποσά από το αρχείο των συναλλαγών. Από τα αποτελέσματα που μας δίνει ο πίνακας (βλ. Εικόνα 4.21), βλέπουμε ότι η ομάδα 1 των καταναλωτών σημειώνει τις υψηλότερες αγορές στο κατάστημα 1 και συγκεκριμένα στις κατηγορίες ΜΗ ΑΛΛΟΙΩΣΙΜΑ και ΦΡΕΣΚΑ ΠΡΟΙΟΝΤΑ. Ακόμη, η ομάδα 3 των καταναλωτών σημειώνει κι αυτή τα υψηλότερα ποσά στο κατάστημα 1. Τα στοιχεία του πίνακα με το μπλε χρώμα αντιπροσωπεύουν τις υψηλότερες τιμές, ενώ με το κόκκινο χρώμα τις χαμηλότερες.

CustomerCode (clusters) 3	1	2	3	Total
Cluster1	30.860,26	10,68		30.870,94
ΜΗ ΑΛΛΟΙΩΣΙΜΑ	18.156,72	6,57		18.163,29
ΦΡΕΣΚΑ ΠΡΟΪΟΝΤΑ	11.332,57			11.332,57
ΟΙΚΙΑΚΟΣ ΕΞΟΠΛΙΣΜΟΣ & ΑΞΕΣΟΥΑΡ	1.103,04	4,11		1.107,15
ΡΟΥΧΙΣΜΟΣ	212,15			212,15
ΗΛΕΚΤΡΟΝΙΚΕΣ & ΗΛΕΚΤΡΙΚΕΣ ΣΥΣΚΕΥΕΣ	55,78			55,78
Cluster3	30.185,57	131,73		30.317,30
ΜΗ ΑΛΛΟΙΩΣΙΜΑ	15.952,55	131,73		16.084,28
ΦΡΕΣΚΑ ΠΡΟΪΟΝΤΑ	12.369,06			12.369,06
ΟΙΚΙΑΚΟΣ ΕΞΟΠΛΙΣΜΟΣ & ΑΞΕΣΟΥΑΡ	1.568,10			1.568,10
ΗΛΕΚΤΡΟΝΙΚΕΣ & ΗΛΕΚΤΡΙΚΕΣ ΣΥΣΚΕΥΕΣ	166,35			166,35
ΡΟΥΧΙΣΜΟΣ	129,51			129,51
Cluster2	7.916,05	2.795,36	192,14	10.903,55
ΜΗ ΑΛΛΟΙΩΣΙΜΑ	4.734,34	1.278,72	107,62	6.120,68
ΦΡΕΣΚΑ ΠΡΟΪΟΝΤΑ	2.769,99	1.324,61	84,52	4.179,12
ΟΙΚΙΑΚΟΣ ΕΞΟΠΛΙΣΜΟΣ & ΑΞΕΣΟΥΑΡ	354,20	192,03		546,23
ΡΟΥΧΙΣΜΟΣ	57,52			57,52
Total	68.961,88	2.937,77	192,14	72.091,79

Εικόνα 4.21

Αφού είδαμε ομαδοποιημένα σε ποια καταστήματα πραγματοποιούν οι καταναλωτές τις αγορές τους, μπορούμε να εντάξουμε και την παράμετρο της συχνότητας. Δηλαδή, να δούμε οι συχνότεροι και οι λιγότερο συχνότεροι (οι οποίοι πραγματοποιούσαν αγορές με υψηλότερα ποσά) πώς μοιράζονται στα τρία καταστήματα και τα συνολικά ποσά των αγορών τους αντίστοιχα.

Θα χρησιμοποιήσουμε ως φίλτρο στον παραπάνω πίνακα (βλ. Εικόνα 4.21), τη συστάδα 3 της Εικόνας 4.15. , η οποία αντιπροσωπεύει τους λιγότερους συχνούς καταναλωτές και τα αποτελέσματα του πίνακα έχουν ως εξής (βλ. Εικόνα 4.22). Παρατηρούμε ότι όλοι πραγματοποιούν τις αγορές τους στο κατάστημα 1, αλλά 74 από αυτούς πραγματοποιούν και στα υπόλοιπα 2 μαγαζιά.

NStores	1	2	3			
CustomerCode (clusters) 3	sumAmount	Count of CustomerCode (clusters)	sumAmount	Count of CustomerCode (clusters)	sumAmount	Count of CustomerCode (clusters)
Cluster3	11.705,83	218				
Cluster1	4.093,66	121				
Cluster2	2.952,70	74	887,00	74	78,36	74
Total	18.752,19	413	887,00	413	78,36	413

Εικόνα 4.22

Στην άλλη περίπτωση, δηλαδή αν επιλέξουμε τη συστάδα 2 με τους συχνότερους (βλ. Εικόνας 4.15), τα αποτελέσματα έχουν ως εξής (βλ. Εικόνα 4.23). Παρατηρούμε πως και οι συχνότεροι καταναλωτές πραγματοποιούν τις αγορές τους στο κατάστημα 1, ενώ μόλις 30 από το σύνολό τους πηγαίνουν στα καταστήματα 2 και 3.

NStores	1	2	3			
CustomerC ode (clusters) 3	sumAmount	Count of CustomerCo de (clusters)	sumAm ount	Count of Customer Code (clusters)	sumAm ount	Count of CustomerC ode (clusters)
Cluster1	8.754,29	510				
Cluster3	1.343,00	30				
Cluster2	456,22	30	196,38	30	19,06	30
Total	10.553,51	570	196,38	570	19,06	570

Εικόνα 4.23

4.3.3 Δημιουργία Αναφορών και χρήση συστάδων

Σε αυτό το σημείο αφού δημιουργήσαμε κάποιες ομάδες προϊόντων και καταναλωτών, οι οποίες ακολουθούν ένα πρότυπο, δηλαδή οι παράμετροι-μεταβλητές που περιλαμβάνουν έχουν στοιχεία που τις διαχωρίζουν, θα δημιουργήσουμε τρεις αναφορές που αφορούν συνολικά την επιχείρηση, το τμήμα πωλήσεων και το τμήμα μάρκετινγκ για δύο περιπτώσεις.

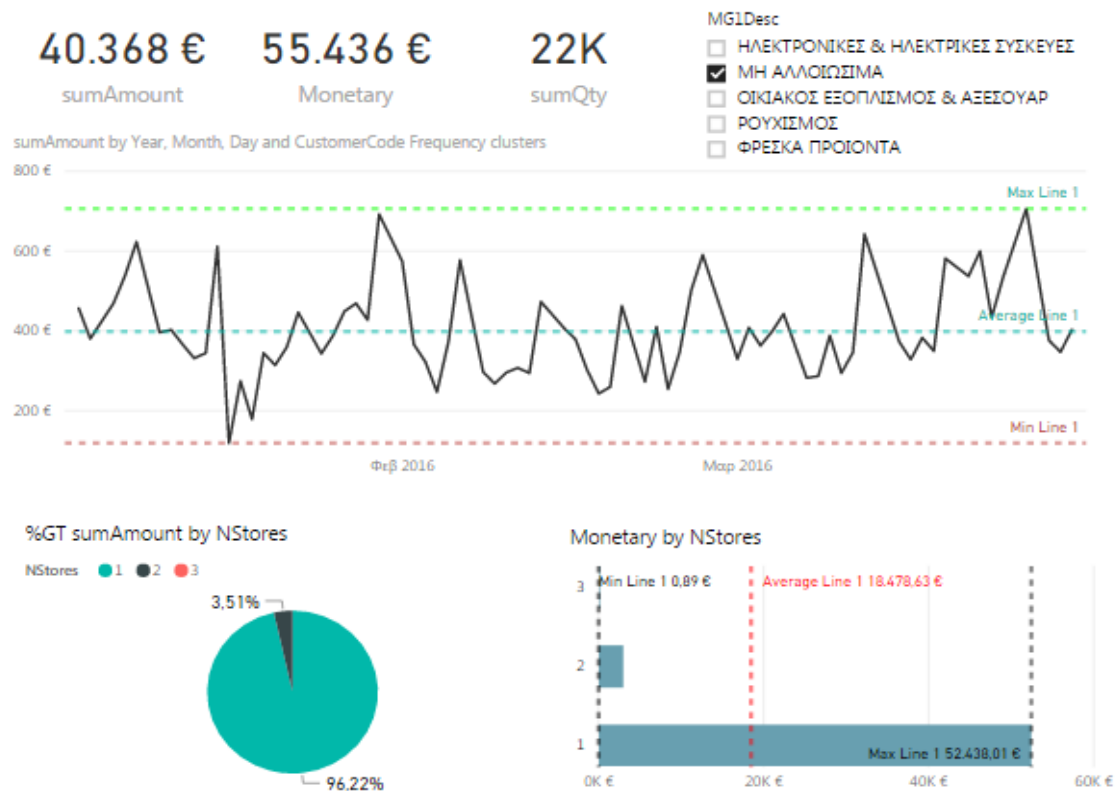
Στην πρώτη περίπτωση θα δημιουργήσουμε αναφορές, στις οποίες δε θα χρησιμοποιήσουμε τις ομάδες-συστάδες που δημιουργήσαμε, ενώ στη δεύτερη περίπτωση θα χρησιμοποιήσουμε τις ομάδες ως φίλτρα ή ως κόφτες, ώστε να εξετάσουμε τις διαφορές των αποτελεσμάτων στις απεικονίσεις μας.

Συνολική Αναφορά

Για τη δημιουργία της αναφοράς που θα αφορά ευρύτερα τα μέλη του οργανισμού, θα χρησιμοποιήσουμε ως απεικονίσεις τις εξής:

- Τρεις κάρτες μοναδικών ροών, οι οποίες θα δείχνουν τα συνολικά ποσά των πωλήσεων, το συνολικό κέρδος και τη συνολική ποσότητα των προϊόντων που πουλήθηκαν.
- Ένα γράφημα γραμμής, το οποίο θα αναπαριστά την πορεία των πωλήσεων για τις ημερήσιες καταγραφές των συναλλαγών και θα περιλαμβάνει τρία αναλυτικά στοιχεία (ελάχιστο, μέγιστο, μέσο όρο).
- Ένα γράφημα πίτας, που θα δείχνει τα ποσοστά των συνολικών πωλήσεων για κάθε κατάσταση.
- Ένα ομαδοποιημένο διάγραμμα ράβδων, το οποίο θα παρουσιάζει τι όφελος έχουμε από κάθε κατάσταση και θα περιλαμβάνει τρία αναλυτικά στοιχεία (ελάχιστο, μέγιστο, μέσο όρο).
- Έναν κόφτη, οποίος θα περιλαμβάνει τις ευρύτερες κατηγορίες των προϊόντων και θα μας διαθέτει την τροποποίηση της αναφοράς σύμφωνα με αυτές τις κατηγορίες.

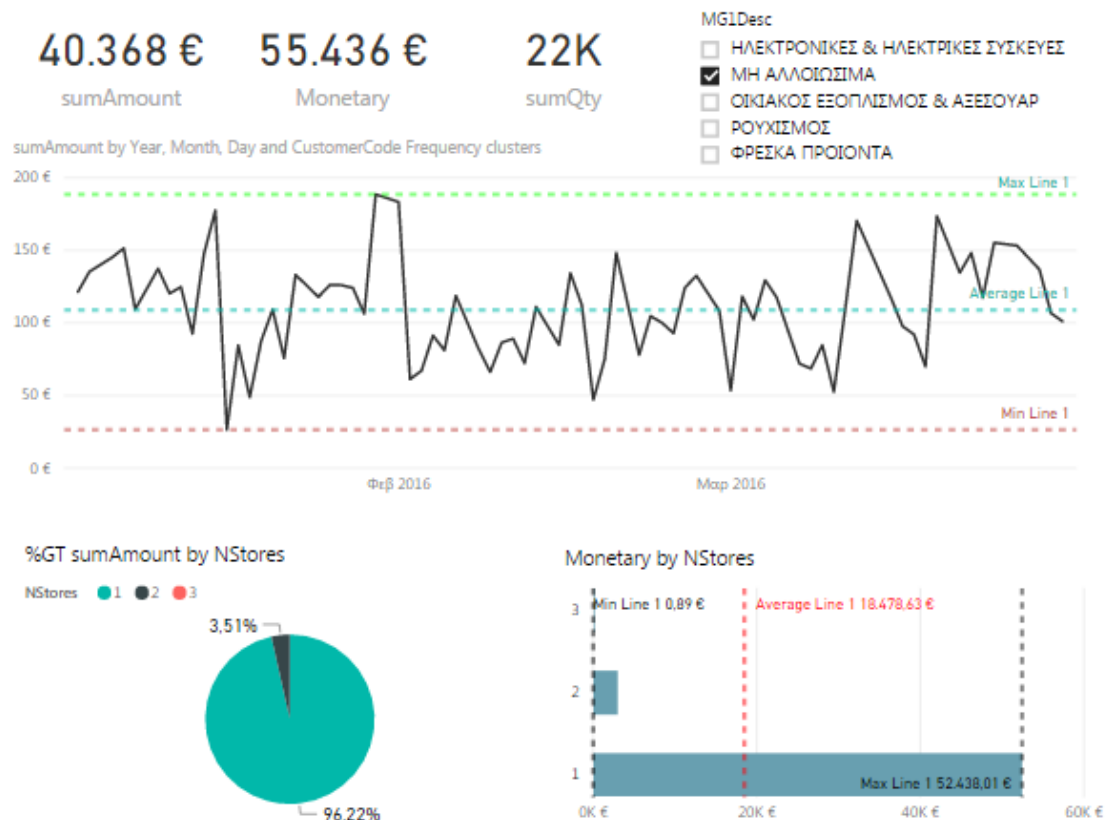
Η μορφή της αναφοράς παρουσιάζεται στην Εικόνα 4.24 κι έχουμε επιλέξει στον κόφτη να μας δίνει τα αποτελέσματα για την κατηγορία των προϊόντων ΜΗ ΑΛΛΟΙΩΣΙΜΑ. Δηλαδή, όλη η αναφορά έχει τροποποιηθεί αυτόματα για αυτήν την κατηγορία. Βλέπουμε τα συνολικά ποσά των πωλήσεών τους, την πορεία τους σύμφωνα με τις ημερήσιες καταγραφές των συναλλαγών. Εδώ να αναφέρουμε ότι έχουμε χρησιμοποιήσει το πεδίο DATE που είχαμε δημιουργήσει κατά την ομαδοποίηση, ώστε να παίρνουμε τις ημερήσιες καταγραφές. Ακόμη, βλέπουμε τα ποσοστά των πωλήσεων για τα συγκεκριμένα προϊόντα και τα οφέλη που έχουμε για κάθε κατάσταση.



Εικόνα 4.24

Ας δούμε όμως τη δεύτερη περίπτωση για τη συγκεκριμένη αναφορά, δηλαδή την τροποποίησή της εάν εφαρμόσουμε τις συστάδες που δημιουργήθηκαν ως φίλτρα.

Έστω ότι επιλέγουμε τις συστάδες 1, 2, 4 της Εικόνας 4.11, τις 1,2, 3 της Εικόνας 4.14 και τις 1, 3 της Εικόνας 4.20. Τις εφαρμόζουμε ως φίλτρα στο γράφημα γραμμής, όπου όπως αναφέραμε αναπαριστά την ημερήσια πορεία σύμφωνα με τις καταγραφές των συναλλαγών. Έχουμε αφαιρέσει δηλαδή ζευγάρια ακραίων τιμών σύμφωνα με τις ομαδοποιήσεις των καταναλωτών. Τα αποτελέσματα της αναφοράς μετά από αυτή την εφαρμογή του φίλτρου φαίνονται στην Εικόνα 4.25.



Εικόνα 4.25

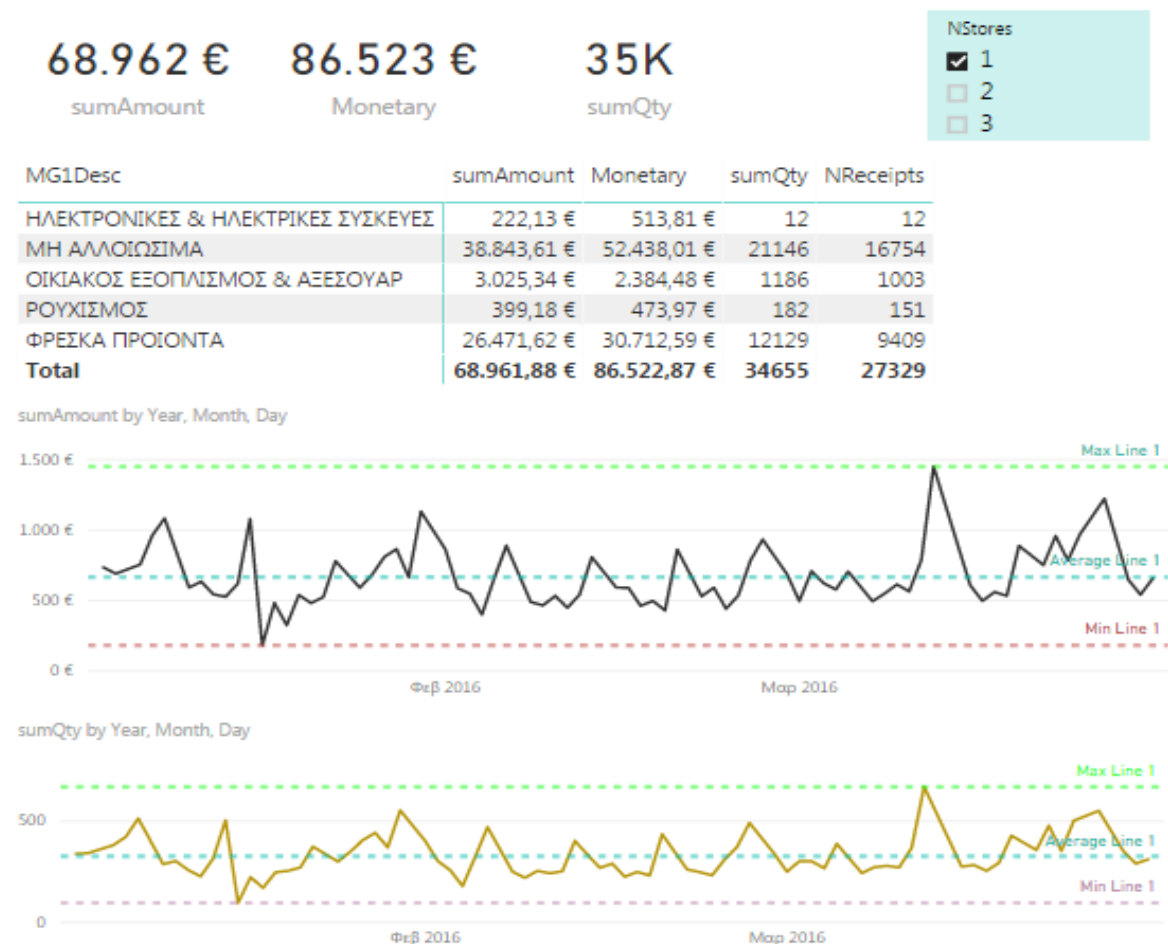
Οι διαφορές στην αναφορά μπορεί να μη φαίνονται μεγάλες, αλλά αυτή η αναφορά αντιπροσωπεύει καλύτερα τα αποτελέσματα της ανάλυσης, διότι το γράφημα γραμμής αναπαριστά ακριβώς την πορεία της συγκεκριμένης κατηγορίας των προϊόντων. Ακόμη, με την αφαίρεση των ομάδων που περιελάμβαναν ακραίες τιμές τα αναλυτικά στοιχεία (ελάχιστο, μέγιστο, μέσος όρος) προσαρμόστηκαν καταλλήλότερα στα δεδομένα μας.

Αναφορά για το Τμήμα Πωλήσεων

Για τη δημιουργία της αναφοράς που θα αφορά τα μέλη του τμήματος των πωλήσεων του οργανισμού, θα χρησιμοποιήσουμε ως απεικονίσεις τις εξής:

- Τρεις κάρτες μοναδικών ροών, οι οποίες θα δείχνουν τα συνολικά ποσά των πωλήσεων, το συνολικό κέρδος και τη συνολική ποσότητα των προϊόντων που πουλήθηκαν.
- Έναν πίνακα matrix, ο οποίος θα δείχνει τις τιμές του συνολικού ποσού των πωλήσεων, το συνολικό κέρδος, τη συνολική ποσότητα που πουλήθηκε και το πλήθος των αποδείξεων που κόπηκαν κατά τις συναλλαγές.
- Ένα γράφημα γραμμής, το οποίο θα αναπαριστά την πορεία των πωλήσεων για τις ημερήσιες καταγραφές των συναλλαγών και θα περιλαμβάνει τρία αναλυτικά στοιχεία (ελάχιστο, μέγιστο, μέσο όρο).
- Ένα γράφημα γραμμής, το οποίο θα αναπαριστά την πορεία των ποσοτήτων των προϊόντων για τις ημερήσιες καταγραφές των συναλλαγών και θα περιλαμβάνει τρία αναλυτικά στοιχεία (ελάχιστο, μέγιστο, μέσο όρο).
- Έναν κόφτη, οποίος θα περιλαμβάνει ως επιλογές τον αριθμό του καταστήματος.

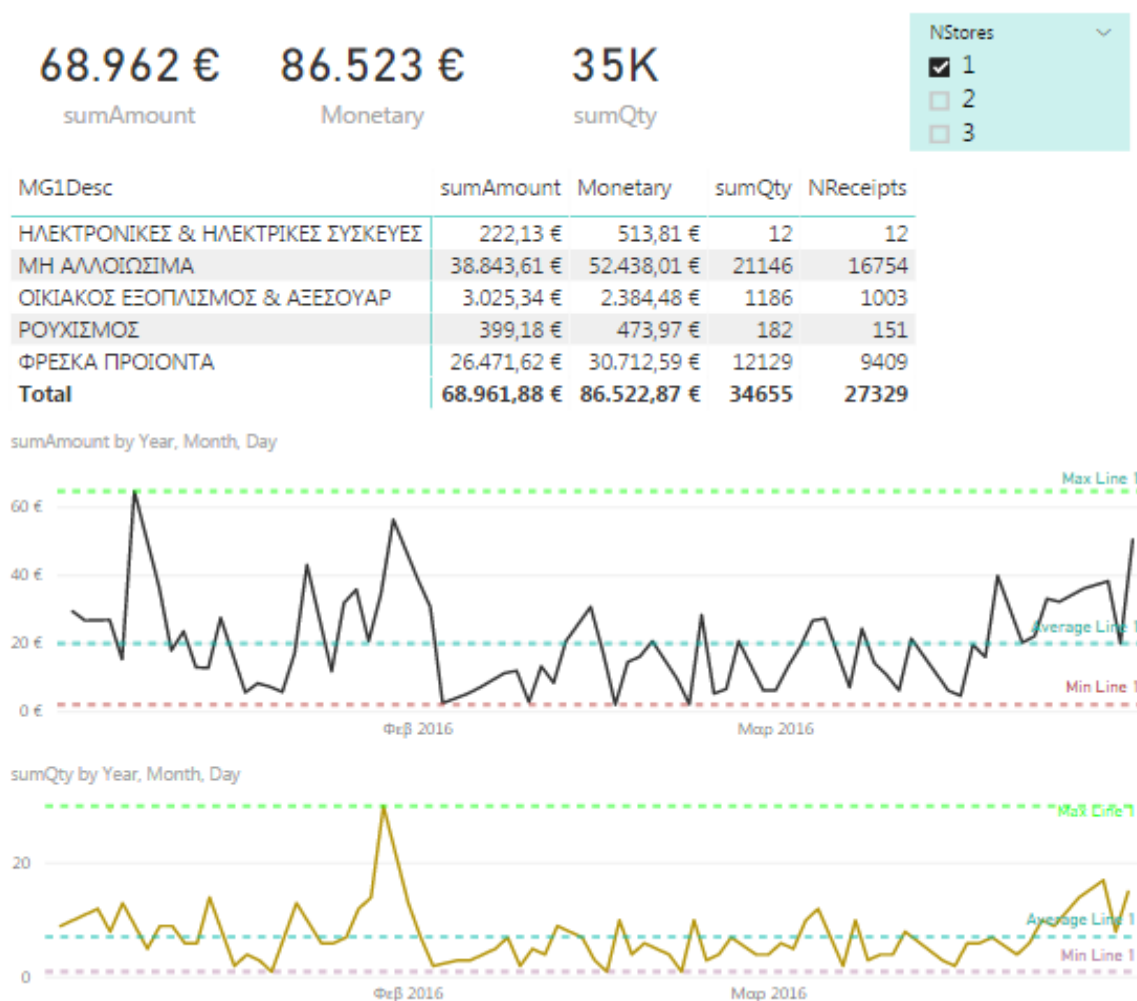
Η μορφή της αναφοράς παρουσιάζεται στην Εικόνα 4.26 κι έχουμε επιλέξει στον κόφτη να μας δίνει τα αποτελέσματα για το ένα κατάστημα. Δηλαδή, όλη η αναφορά έχει τροποποιηθεί αυτόματα για το συγκεκριμένο κατάστημα. Βλέπουμε τα συνολικά ποσά των πωλήσεών του, το συνολικό κέρδος, τη συνολική ποσότητα, πώς μοιράζονται στις ευρύτερες κατηγορίες των προϊόντων μας και την πορεία των πωλήσεών του σύμφωνα με τις ημερήσιες καταγραφές των συναλλαγών. Ακόμη, βλέπουμε την πορεία των συνολικών ποσοτήτων για το συγκεκριμένο κατάστημα σύμφωνα με τις ημερήσιες καταγραφές των συναλλαγών.



Εικόνα 4.26

Ας δούμε όμως τη δεύτερη περίπτωση για τη συγκεκριμένη αναφορά, δηλαδή την τροποποίησή της εάν εφαρμόσουμε τις συστάδες που δημιουργήθηκαν ως φίλτρα.

Έστω ότι επιλέγουμε τις συστάδες 1 της Εικόνας 4.1, τις 1, 3 της Εικόνας 4.5 και την 1 της Εικόνας 4.7. Τις εφαρμόζουμε ως φίλτρα στα γραφήματα γραμμής, όπου όπως αναφέραμε αναπαριστούν την ημερήσια πορεία σύμφωνα με τις καταγραφές των συναλλαγών. Έχουμε αφαιρέσει δηλαδή τα ζευγάρια ακραίων τιμών σύμφωνα με τις ομαδοποιήσεις των προϊόντων. Τα αποτελέσματα της αναφοράς μετά από αυτή την εφαρμογή του φίλτρου φαίνονται στην Εικόνα 4.27.



Εικόνα 4.27

Οι διαφορές στην αναφορά είναι μεγάλες που σημαίνει πως κατά την καταγραφή των συναλλαγών έγιναν αρκετά λάθη ή υπήρχαν αρκετά κενά που διαστρέβλωναν τα αποτελέσματα. Κατά την ομαδοποίηση των προϊόντων αφαιρέθηκαν αρκετά, αλλά σαφώς αφαιρέθηκε και ένα μέρος πληροφοριών. Εάν τις εφαρμόζαμε σε έναν πίνακα απεικόνισης θα παίρναμε τελείως διαφορετικά αποτελέσματα και λανθασμένα. Και σε αυτήν την περίπτωση όμως, τα γραφήματα γραμμής αναπαριστούν καλύτερα την πορεία του συγκεκριμένου καταστήματος και η προσαρμογή των αναλυτικών στοιχείων (ελάχιστο, μέγιστο, μέσος όρος) είναι καταλληλότερη.

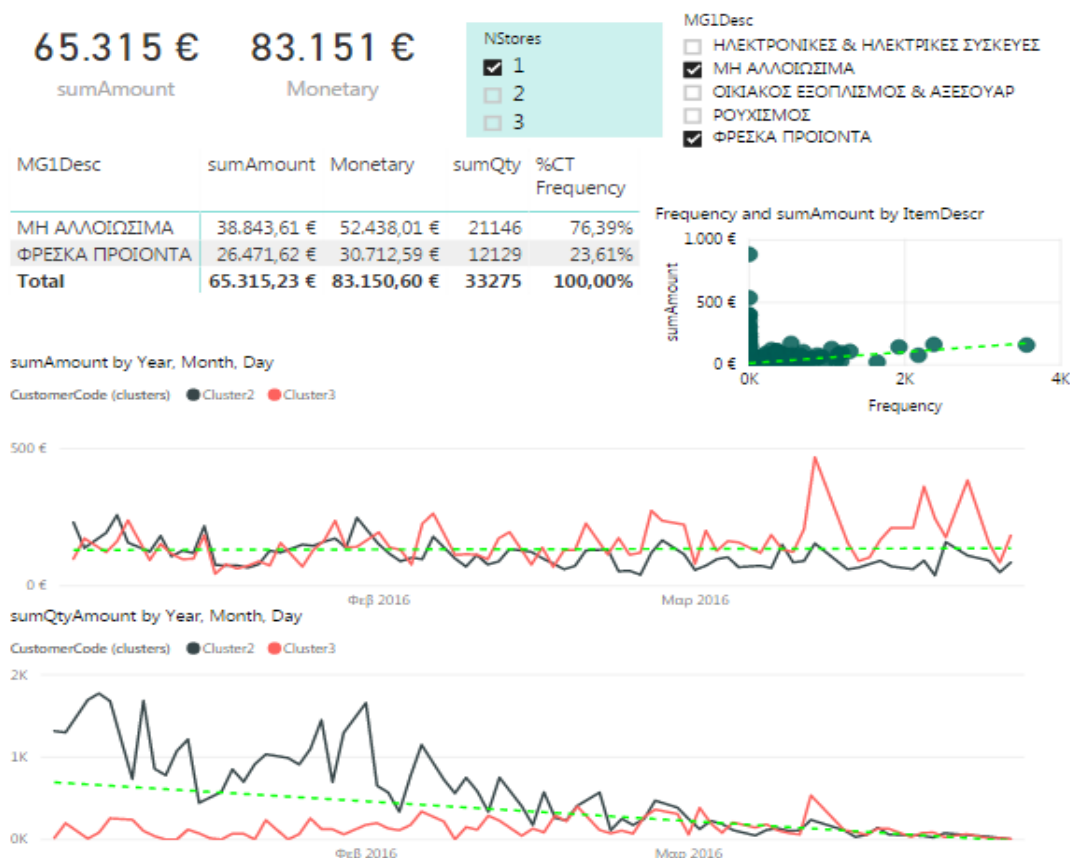
Αναφορά για το Τμήμα Μάρκετινγκ

Για τη δημιουργία της αναφοράς που θα αφορά τα μέλη του τμήματος του μάρκετινγκ του οργανισμού, θα χρησιμοποιήσουμε ως απεικονίσεις τις εξής:

- Δύο κάρτες μοναδικών ροών, οι οποίες θα δείχνουν τα συνολικά ποσά των πωλήσεων και το συνολικό κέρδος των προϊόντων που πουλήθηκαν.
- Έναν πίνακα matrix, ο οποίος θα δείχνει τις τιμές του συνολικού ποσού των πωλήσεων, το συνολικό κέρδος, τη συνολική ποσότητα που πουλήθηκε και το συνολικό ποσοστό της συχνότητας που πραγματοποιήθηκαν οι πωλήσεις σύμφωνα με τις συναλλαγές.

- Ένα διάγραμμα διασποράς, το οποίο θα αναπαριστά τη σχέση του συνολικού ποσού των πωλήσεων με τη συχνότητα που πραγματοποιήθηκαν. Ακόμη θα περιλαμβάνει μία αναλυτική γραμμή που θα δείχνει τη συνολική τάση σύμφωνα με τις συναλλαγές.
- Ένα γράφημα γραμμής, το οποίο θα αναπαριστά την πορεία των πωλήσεων για τους συχνούς (cluster 2) και τους λιγότερους συχνούς (cluster 3) καταναλωτές σύμφωνα με τις ημερήσιες καταγραφές των συναλλαγών. Επίσης, θα περιλαμβάνει μία αναλυτική γραμμή που θα δείχνει τη συνολική τους τάση σύμφωνα με τις συναλλαγές
- Ένα γράφημα γραμμής, το οποίο θα αναπαριστά την πορεία των ποσοτήτων των προϊόντων που αγοράστηκαν από τους συχνούς (cluster 2) και τους λιγότερους συχνούς (cluster 3) καταναλωτές σύμφωνα με τις ημερήσιες καταγραφές των συναλλαγών. Επίσης, θα περιλαμβάνει μία αναλυτική γραμμή που θα δείχνει τη συνολική τους τάση σύμφωνα με τις συναλλαγές
- Έναν κόφτη, οποίος θα περιλαμβάνει ως επιλογές τον αριθμό του καταστήματος.
- Έναν κόφτη, οποίος θα περιλαμβάνει τις ευρύτερες κατηγορίες των προϊόντων.

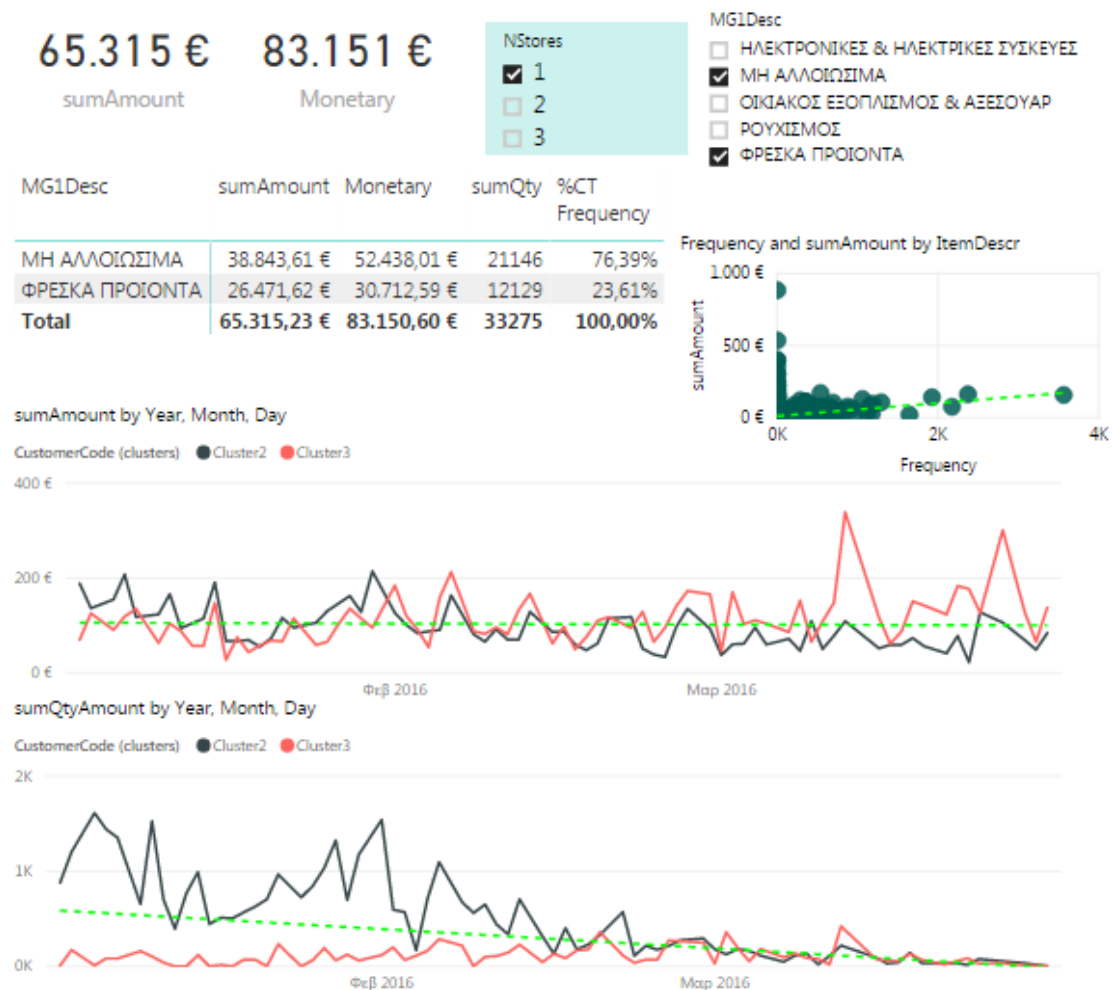
Η μορφή της αναφοράς παρουσιάζεται στην Εικόνα 4.28 κι έχουμε επιλέξει στους κόφτες να μας δίνει τα αποτελέσματα για το ένα κατάστημα και για τις κατηγορίες ΜΗ ΑΛΛΟΙΩΣΙΜΑ και ΦΡΕΣΚΑ ΠΡΟΙΟΝΤΑ, που κατά τη δημιουργία των ομαδοποιήσεων και με τη χρήση απεικονίσεων είδαμε ότι έχουν τα υψηλότερα ποσά των πωλήσεων. Ακόμη βλέπουμε τα συνολικά ποσά των πωλήσεών τους, το συνολικό τους κέρδος, τη συνολική ποσότητά των προϊόντων που αγοράστηκαν και το συνολικό ποσοστό της συχνότητας των πωλήσεών τους για τις επιλογές των κοφτών. Επιπλέον, βλέπουμε την πορεία των συνολικών πωλήσεων και των συνολικών ποσοτήτων των προϊόντων σύμφωνα με τις ημερήσιες καταγραφές των συναλλαγών.



Εικόνα 4.28

Ας δούμε όμως τη δεύτερη περίπτωση για τη συγκεκριμένη αναφορά, δηλαδή την τροποποίησή της εάν εφαρμόσουμε τις συστάδες που δημιουργήθηκαν ως φίλτρα.

Έστω ότι επιλέγουμε τις συστάδες 1, 3 της Εικόνας 4.20 και τις 1, 3 της Εικόνας 4.5. Τις εφαρμόζουμε ως φίλτρα στα γραφήματα γραμμής, όπου όπως αναφέραμε αναπαριστούν την ημερήσια πορεία σύμφωνα με τις καταγραφές των συναλλαγών. Έχουμε προσθέσει δηλαδή στα γραφήματα δύο ομάδες που αφορούν τους καταναλωτές και δύο ομάδες που αφορούν τα προϊόντα. Ο λόγος που δημιουργήθηκαν παραπάνω αυτές οι ομάδες είναι διότι περιλαμβάνουν κάποια κοινά χαρακτηριστικά-πρότυπα, τα οποία θέλουμε να περιλαμβάνονται στις απεικονίσεις μας. Αυτά τα χαρακτηριστικά προσδίδουν παραπάνω πληροφορία στα γραφήματα, διότι με αυτόν τον τρόπο προσθέτονται περιπτώσεις που ίσως με την παράλειψή τους να είχαμε αποτελέσματα λιγότερης ακρίβειας. Τα αποτελέσματα της αναφοράς μετά από αυτή την εφαρμογή των συστάδων του φίλτρα φαίνονται στην Εικόνα 4.29.



Εικόνα 4.29

Οι διαφορές στην αναφορά δεν είναι μεγάλες που σημαίνει πως οι ομάδες που προσθέσαμε περιλαμβάνονταν στα γραφήματα γραμμών. Υπάρχουν όμως κάποια σημεία που δεν υπήρχαν κι αυτά τα σημεία προσδίδουν τα πρότυπα που περιλαμβάνουν αυτές οι ομάδες. Σε μία ημερήσια εξέταση από ένα μέλος του τμήματος μάρκετινγκ μπορεί να δώσει άλλα αποτελέσματα σε σύγκριση με την προηγούμενη αναφορά. Ακόμη, παρατηρούμε πως δεν επηρεάστηκαν οι αναλυτικές γραμμές που δείχνουν τις τάσεις, διότι οι συγκεκριμένες πληροφορίες που προστέθηκαν από τις ομάδες δεν περιελάμβαναν ακραίες τιμές. Θα μπορούσαμε να είχαμε τεράστιες διαφορές σε σύγκριση με την προηγούμενη αναφορά και θα αποτελούσε ένα σημείο έρευνας από το τμήμα μάρκετινγκ με την προϋπόθεση ότι δεν αποτελούσε λάθος της καταγραφής των συναλλαγών. Επιπλέον, εάν εφαρμόζαμε αυτές τις ομάδες ως φίλτρα σε έναν πίνακα απεικόνισης θα παίρναμε τελείως διαφορετικά αποτελέσματα και μη συνολικά, διότι αποτελούν δείγματα των δεδομένων.

Τέλος, και σε αυτήν την περίπτωση της αναφοράς, τα γραφήματα γραμμής αναπαριστούν καλύτερα την πορεία του συγκεκριμένου καταστήματος στις συγκεκριμένες κατηγορίες προϊόντων.

4.4 ΣΥΜΠΕΡΑΣΜΑΤΑ

Το Power BI Desktop μας διαθέτει πολλές δυνατότητες ώστε να εξετάσουμε τα δεδομένα μας. Η πληθώρα των επιλογών στις απεικονίσεις μας επιτρέπει την παρουσίαση των αποτελεσμάτων μας από πολλές σκοπιές.

Όσον αφορά τη διαδικασία συσταδοποίησης που πραγματοποιεί το πρόγραμμα είδαμε πως μας διαθέτει δύο επιλογές, την αυτόματη με την οποία το ίδιο το πρόγραμμα αποφασίζει για το πλήθος των ομάδων που θα δημιουργήσει και την επιλογή όπου ο χρήστης επιλέγει το πλήθος των ομάδων και το πρόγραμμα τις δημιουργεί.

Εφαρμόσαμε και τις δύο επιλογές όπου είδαμε πως σε μερικές περιπτώσεις που επιθυμούμε ένα πιο συγκεκριμένο μέρος των δεδομένων μας είναι καλύτερο να επιλέγουμε εμείς το πλήθος των ομάδων. Ακόμη, όταν υπάρχουν αρκετές μεταβλητές με ακραίες τιμές μπορούμε με τη χρήση της συσταδοποίησης να τις αφαιρέσουμε από τα αποτελέσματα της ανάλυσής μας.

Αφού ομαδοποιήσαμε τα δεδομένα μας δημιουργήσαμε αναφορές και χρησιμοποιήσαμε μερικές από τις ομάδες ως φίλτρα στις απεικονίσεις τους. Είδαμε πως μετά την εφαρμογή των φίλτρων τα αποτελέσματά μας είχαν επηρεαστεί. Αυτό οφείλεται στο ότι επιλέξαμε συγκεκριμένες συστάδες όπου δεν περιελάμβαναν ακραίες τιμές. Έτσι, βελτιώσαμε τα αποτελέσματα των απεικονίσεών μας και κατά συνέπεια την ακρίβειά τους. Ακόμη, οι επιλογές μας στις συστάδες δεν έγιναν αυθαίρετα, αλλά ανάλογα με τα πεδία των απεικονίσεών μας.

Τέλος, η δημιουργία των συστάδων μας βοηθά στο να εξετάσουμε ένα μέρος των δεδομένων μας όπου παρουσιάζει κοινή συμπεριφορά, δηλαδή ένα πιο στοχευόμενο τμήμα των δεδομένων μας, όπως οι συχνότεροι πελάτες που είδαμε στην αναφορά για το τμήμα του μάρκετινγκ.

ΚΕΦΑΛΑΙΟ 5:

ΣΥΜΠΕΡΑΣΜΑΤΑ

Τα εργαλεία και οι τεχνικές της Επιχειρηματικής Ευφυΐας μπορούν να παρέχουν πολλά οφέλη στους οργανισμούς, διότι τους βοηθά να αντλούν πολύτιμες πληροφορίες από τα δεδομένα που διατηρούν. Η διαχείριση και η ανάλυση των δεδομένων έχει αποσπάσει σημαντικά το ενδιαφέρον πολλών οργανισμών, διότι όχι μόνο μπορεί να βελτιώσει την οργάνωση κι ευρύτερα τις διαδικασίες τους, αλλά μπορεί να μετατρέψει τα δεδομένα τους σε κερδοφόρα έσοδα. Ακόμη, η αναγνώριση των κερδοφόρων προτύπων σε έναν οργανισμό, όπως οι συχνότεροι πελάτες ή το προϊόν που ενδυναμώνει τη σχέση τους με τους πελάτες μπορεί να οδηγήσει στην βελτίωση της αποδοτικότητας και της αποτελεσματικότητας όχι μόνο των λειτουργιών ενός οργανισμού, αλλά της εικόνας του γενικότερα.

Μέσα στις δυνατότητες που διαθέτουν τα εργαλεία της Επιχειρηματικής Ευφυΐας είναι και η κατανόηση της αγοράς που απευθύνεται ένας οργανισμός. Μέσω της ανάλυσης πηγάζουν πολλές πιθανές ευκαιρίες για επιχειρηματικές πρωτοβουλίες κι αποφάσεις, όπου με την ανίχνευσή τους διατηρούν τα μέλη ενός οργανισμού σε ετοιμότητα. Επιπλέον, η κατανόηση των καταναλωτών και η ανίχνευση των προτύπων που ακολουθούν αποτελεί θεμέλιο για την βέλτιστη εξυπηρέτησή τους. Η κατάλληλη διαχείριση κι ανάλυση αυτών των προτύπων αποτελεί βασικό παράγοντα στην εκπλήρωση των στόχων ενός οργανισμού, όπως την επίτευξη του ανταγωνιστικού πλεονεκτήματος.

Ένα σημαντικό μέρος της ανάλυσης των επιχειρηματικών δεδομένων αποτελεί η ενημέρωση των τμημάτων ενός οργανισμού με τη δημιουργία αναφορών. Υπάρχουν πολλές προτάσεις για τις μεθόδους δημιουργίας τους και τη δομή τους, αλλά τα αποτελέσματά που παρουσιάζουν κι οι μορφές των απεικονίσεων που χρησιμοποιούνται εξαρτώνται σημαντικά από την πλευρά και ιδιαίτερα από την κρίση του αναλυτή.

Τα συστήματα Επιχειρηματικής Ευφυΐας διαθέτουν πολλά εργαλεία που αποσκοπούν στην βαθύτερη επεξεργασία-τροποποίηση των δεδομένων, ώστε να βελτιώσουν την “αξία” που εγκυμονούν, η οποία δεν είναι εφικτό να διακριθεί από τα μέλη μιας οργάνωσης. Η χρήση των αναφορών έχει ως στόχο την ενημέρωση αυτών των μελών μέσω της παρουσίασης των αποτελεσμάτων μιας επιχειρηματικής ανάλυσης χρησιμοποιώντας διάφορες μορφές απεικονίσεων. Επιπλέον, οι παρουσιάσεις τέτοιων αναφορών μπορούν να κρατούν ενήμερα τα μέλη κάθε τμήματος ενός οργανισμού για την κατάστασή του ευρύτερα.

Στην παρούσα εργασία, πραγματοποιήσαμε την ανάλυση κάποιων επιχειρηματικών δεδομένων με τη χρήση του εργαλείου Power BI Desktop. Αρχικά, χρησιμοποιήσαμε τη μέθοδο της συσταδοποίησης των δεδομένων που διαθέτει το πρόγραμμα για να Στην παρούσα εργασία, πραγματοποιήσαμε την ανάλυση κάποιων επιχειρηματικών

δεδομένων με τη χρήση του εργαλείου Power BI Desktop. Αρχικά, χρησιμοποιήσαμε τη μέθοδο της συσταδοποίησης των δεδομένων που διαθέτει το πρόγραμμα για να ομαδοποιήσουμε τα προϊόντα και τους καταναλωτές μας σύμφωνα με τις υπόλοιπες μεταβλητές που περιελάμβανε το αρχείο των δεδομένων μας. Εφαρμόσαμε τη μέθοδο σε διάφορα “ζευγάρια” των δεδομένων μας με σκοπό να ανιχνεύσουμε πιθανά πρότυπα που υπήρχαν στα δεδομένα μας. Στη συνέχεια, χρησιμοποιήσαμε τα πεδία από τις ομάδες που δημιουργήθηκαν ως φίλτρα σε διάφορες μορφές απεικονίσεων μας ώστε να δίνουν πιο καθορισμένα αποτελέσματα. Παρατηρήσαμε πως τα αποτελέσματα των μορφών απεικονίσεων με τη χρήση ή μη των συστάδων ως φίλτρα είχαν σημαντική διαφορά. Αυτό σημαίνει πως οι τάσεις και τα πρότυπα που “κρύβονται” στα δεδομένα μας έχουν ισχυρή επιρροή στα αποτελέσματα και κατά συνέπεια στα ορθά συμπεράσματά μας. Ακόμη, δεν ακολουθήσαμε τη συνολική ομαδοποίηση των δεδομένων που αφορά το προϊόν και ξεχωριστά τον καταναλωτή, αλλά εφαρμόσαμε σταδιακά τη δημιουργία ομάδων και για τις δύο μεταβλητές μας. Στην πορεία της ανάλυσης, αποκλείαμε τις ομάδες που κατά την κρίση μας δεν ήταν τόσο σημαντικές ή δεν έδειχναν να έχουν κάποια τάση-διαφορετικότητα και χρησιμοποιούσαμε τις υπόλοιπες ως φίλτρα στις μορφές των απεικονίσεων μας για να λαμβάνουμε υπόψη τις τάσεις τους και να έχουμε πιο προσδιορισμένα στοιχεία στα αποτελέσματά μας.

Στη συνέχεια, δημιουργήσαμε τρεις αναφορές για δύο περιπτώσεις. Η μία περίπτωση δεν περιελάμβανε τη χρήση συστάδων ως φίλτρα, ενώ η δεύτερη περιελάμβανε. Στην πρώτη αναφορά, που περιελάμβανε στοιχεία για την ευρύτερη εικόνα της επιχείρησης, μετά την εφαρμογή των φίλτρων δεν υπήρχαν μεγάλες διαφορές. Στην δεύτερη αναφορά, που περιελάμβανε στοιχεία για το τμήμα των πωλήσεων, μετά την εφαρμογή των φίλτρων είχε μεγάλες διαφορές λόγω της ύπαρξης λαθών στα δεδομένα μας. Στην τρίτη αναφορά, που περιείχε στοιχεία για το τμήμα μάρκετινγκ, μετά την εφαρμογή των φίλτρων είχε μικρές διαφορές. Και στις τρεις αναφορές όμως προστέθηκαν πληροφορίες που μπορεί να μην διακρίνονται στις μορφές απεικονίσεων, αλλά επηρεάζουν σε ένα βαθμό τα αποτελέσματα μιας ανάλυσης. Οι νέες πληροφορίες δεν είναι απόλυτα σωστές και χρειάζονται εξέταση από τα αρμόδια μέλη, αλλά η παράλειψη κάποιων παρατηρήσεων που προσδίδουν χαρακτηριστικά στα δεδομένα μας μπορεί να οδηγήσει μέχρι και σε μία λανθασμένη επιχειρηματική απόφαση.

Όσον αφορά τη δημιουργία των αναφορών, είναι απαραίτητο να αναφερθεί πως οι ιεραρχίες που δημιουργούμε στα τα δεδομένα μας και η χρήση τους στις αναφορές μας προσδίδουν πολλαπλά επίπεδα στα αποτελέσματα της ανάλυσής μας. Μας διαθέτουν την επιλογή μιας πιο καθορισμένης προβολής σύμφωνα με το τι αναζητάμε. Ακόμη, η χρήση των αναλυτικών γραμμών μπορεί να δώσει όχι μόνο τα σημαντικά σημεία ενός γραφήματος, αλλά τα ίδια τα συμπεράσματα από τα αποτελέσματά μας. Επιπλέον, η χρήση συστάδων είτε ως φίλτρα είτε ως κόφτες μπορούν να φτάσουν το επίπεδο της ανάλυσης και κατά συνέπεια των αποτελεσμάτων σε σημαντικά υψηλότερο βαθμό. Βέβαια, οι τάσεις και τα πρότυπα που πηγάζουν από τα δεδομένα δεν είναι απόλυτα και δε μπορούν να θεωρηθούν

μόνιμα για το επόμενο χρονικό διάστημα. Για παράδειγμα, η εφαρμογή μιας εκστρατείας μάρκετινγκ μπορεί να επηρεάσει κατά πολύ μεγάλο βαθμό τις συστάδες των πελατών.

Τα ανοιχτά ζητήματα που θα μπορούσαν να διερευνηθούν από μια μελλοντική επέκταση της παρούσας εργασίας είναι σε τι βαθμό είναι ερμηνεύσιμες οι αναφορές από τα μέλη ενός οργανισμού είτε περιλαμβάνουν τις τάσεις, τα πρότυπα ή τα αναλυτικά στοιχεία είτε όχι. Ακόμη, θα μπορούσε μια επιχειρηματική απόφαση ή ευκαιρία να στηριχτεί σε μία ομάδα-συστάδα των αποτελεσμάτων μιας ανάλυσης; Και αν ναι σε τι βαθμό και για ποιο χρονικό διάστημα;

BIBΛΙΟΓΡΑΦΙΑ

- [1] Improving Decision Quality: The Role of Business Intelligence Lucian L. Visinescu^a, Mary C. Jones^b, and Anna Sidorova^a ^aTexas State University, San Marcos, TX, USA; ^bUniversity of North Texas, Denton, TX, USA
- [2] Jayanthi Ranjan, (2008) "Business justification with business intelligence", VINE, Vol. 38 Issue: 4, pp.461-475, <https://doi.org/10.1108/03055720810917714>
- [3] C.K.M. Lee^a, H.C.W. Lau^b, G.T.S. Hob^c, William Hoc^c, (2009) "Design and development of agent-based procurement system to enhance business intelligence"
- [4] Giovinazzo, W. A. (2002). Internet-enabled business intelligence. USA: Prentice Hall.
- [5] Lustig, I., Dietrich, B., Johnson, C., & Dziekan, C. (2010). The analytics journey. *Analytics Magazine*, 3(6), 11-13.
- [6] Gandomi, A., & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics. *International Journal of Information Management*, 35(2), 137-144.
- [7] Gröger, C., Schwarz, H., & Mitschang, B. (2014, May). Prescriptive analytics for recommendation-based business process optimization. In *International Conference on Business Information Systems* (pp. 25-37). Springer, Cham.
- [8] Evans, J. R., & Lindner, C. H. (2012). Business analytics: The next frontier for decision sciences. *Decision Line*, 43(2), 4-6.
- [9] Borut Hocevar, Jurij Jaklic, (2010) "Assesing Benefits of Business Intelligence Systems- A Case Study", Management, Vol. 15, pp.87-119
- [10] Gandomi, A & Haider, M. (2015). Beyond the hype: Big data concepts, methods, and analytics, *International Journal of Information Management*, 35(2), 137-144.
- [11] Siegel, E. (2013). Predictive Analytics: The Power to Predict Who Will Click, Buy, Lie, or Die. Wiley, 2013.
- [12] Nofal, M. I., & Yusof, Z. M. (2013). Integration of business intelligence and enterprise resource planning within organizations. *Procedia Technology*, 11, 658-665.
- [13] Melchert, F., Winter, R., & Klesse, M. (2004, August). Aligning process automation and business intelligence to support corporate performance management. Association for Information Systems.
- [14] Elbashir, M. Z., Collier, P. A., & Davern, M. J. (2008). Measuring the effects of business intelligence systems: The relationship between business process and organizational performance. *International Journal of Accounting Information Systems*, 9(3), 135-153.
- [15] Ramakrishnan, T., Jones, M. C., & Sidorova, A. (2012). Factors influencing business intelligence (BI) data collection strategies: An empirical investigation. *Decision Support Systems*, 52(2), 486-496.
- [16] Trkman, P., McCormack, K., De Oliveira, M. P. V., & Ladeira, M. B. (2010). The impact of business analytics on supply chain performance. *Decision Support Systems*, 49(3), 318-327.

- [17] Citroen, C. L. (2011). The role of information in strategic decision-making. *International Journal of Information Management*, 31(6), 493-501.
- [18] Ahmed, S. R. (2004, April). Applications of data mining in retail business. In *Information Technology: Coding and Computing, 2004. Proceedings. ITCC 2004. International Conference on* (Vol. 2, pp. 455-459). IEEE.
- [19] Huang, Z. (1997). A fast clustering algorithm to cluster very large categorical data sets in data mining. *DMKD*, 3(8), 34-39.
- [20] Sahama, T. R., & Croll, P. R. (2007, January). A data warehouse architecture for clinical data warehousing. In *Proceedings of the fifth Australasian symposium on ACSW frontiers-Volume 68* (pp. 227-232). Australian Computer Society, Inc..
- [21] Agarwal, R., & Srikant, R. (1994, September). Fast algorithms for mining association rules. In *Proc. of the 20th VLDB Conference* (pp. 487-499).
- [22] Baker, R. S. J. D. (2010). Data mining for education. *International encyclopedia of education*, 7(3), 112-118.
- [23] Larose, D. T. (2005). *Discovering Knowledge in Data: An Introduction to Data Mining* John Wiley & Sons. Inc., New Jersey.
- [24] Chaudhuri, S., & Dayal, U. (1997). An overview of data warehousing and OLAP technology. *ACM Sigmod record*, 26(1), 65-74.
- [25] Işık, Ö., Jones, M. C., & Sidorova, A. (2013). Business intelligence success: The roles of BI capabilities and decision environments. *Information & Management*, 50(1), 13-23.
- [26] Kumari, M., & Godara, S. (2011). Comparative study of data mining classification methods in cardiovascular disease prediction 1.
- [27] Liu, B., Zhao, K., Benkler, J., & Xiao, W. (2006, August). Rule interestingness analysis using OLAP operations. In *Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 297-306). ACM.
- [28] MacQueen, J. (1967, June). Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability* (Vol. 1, No. 14, pp. 281-297).
- [29] Negash, S. (2004). Business intelligence. *The communications of the Association for Information Systems*, 13(1), 54.
- [30] Yeung, K. Y., Fraley, C., Murua, A., Raftery, A. E., & Ruzzo, W. L. (2001). Model-based clustering and data transformations for gene expression data. *Bioinformatics*, 17(10), 977-987.
- [31] Yin, Y., Kaku, I., Tang, J., & Zhu, J. (2011). *Data mining: Concepts, methods and applications in management and engineering design*. Springer Science & Business Media.

WEB SITES

[1] <https://powerbi.microsoft.com>