



ΠΟΛΥΤΕΧΝΕΙΟ ΚΡΗΤΗΣ

ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ ΠΑΡΑΓΩΓΗΣ ΚΑΙ ΔΙΟΙΚΗΣΗΣ

ΕΠΙΧΕΙΡΗΜΑΤΙΚΗ ΕΥΦΥΪΑ ΚΑΙ ΣΧΕΣΕΙΣ ΕΤΑΙΡΕΙΑΣ
ΜΕ ΤΟ ΚΑΤΑΝΑΛΩΤΙΚΟ ΤΗΣ ΚΟΙΝΟ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΔΕΜΕΣΤΙΧΑΣ Ν. ΜΙΧΑΗΛ

ΕΠΙΒΛΕΠΟΝΤΕΣ ΚΑΘΗΓΗΤΕΣ:

ΚΑΘ. ΜΟΥΣΤΑΚΗΣ ΒΑΣΙΛΕΙΟΣ

ΧΑΝΙΑ, ΣΕΠΤΕΜΒΡΙΟΣ 2022

Περίληψη

Στον κόσμο των επιχειρήσεων, η σύγχρονη εποχή χαρακτηρίζεται σε μεγάλο βαθμό από τη δυνατότητα συλλογής και διαχείρισης δεδομένων, τόσο για τη θέση των επιχειρήσεων απέναντι στον ανταγωνισμό, όσο και για την κερδοφόρα τους ανάπτυξη.

Η σωστή και μεθοδική ανάλυση δεδομένων (data) οδηγεί στη λήψη πολύτιμων πληροφοριών, που στόχο έχουν την κατάρτιση στρατηγικών σχεδίων και σχετίζονται με την διαχείριση κρίσεων, με τον έλεγχο και την ελαχιστοποίηση του ρίσκου και οπτικοποιούν με απλότητα και σαφήνεια δυσνόητα και πολύπλοκα συμπεράσματα.

Νέες τεχνολογίες λογισμικών στον χώρο της Επιχειρηματικής Ευφυΐας καταστούν διαχειρίσιμους τους τεράστιους όγκους δεδομένων σε προσπελάσιμες βάσεις, με αποτέλεσμα η λήψη έγκυρων αποφάσεων να πραγματοποιείται ορθολογικά και εύστοχα.

Στην παρούσα εργασία επεξηγούνται οι όροι Μεγάλα Δεδομένα (Big Data), Ανοιχτά και Δημόσια Δεδομένα, οι πηγές και οι μορφές τους, περιγράφονται τα τυπικά συστήματα Επιχειρηματικής Ευφυΐας και ορισμένες βασικές μέθοδοι εξόρυξης πληροφορίας και γνώσης.

Τέλος, με τη χρήση του λογισμικού (Tableau) παράγονται αναφορές, οπτικοποιούνται δεδομένα και αναδεικνύονται περαιτέρω οι δυνατότητες και η ευελιξία που προσφέρει ένα τέτοιο εργαλείο στο χώρο του επιχειρείν.

Ευχαριστίες

Θα ήθελα να εκφράσω τις ευχαριστίες σε όσους συνέβαλαν ώστε να ολοκληρωθεί η εργασία μου και γενικότερα οι σπουδές μου. Κατ' αρχάς, θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου, τον κ. Μουστάκη Βασίλειο που μου έδωσε την ευκαιρία να εργαστώ πάνω σε έναν τόσο ενδιαφέρον τομέα και την καθοδήγηση που μου έδειξε όλο αυτό τον καιρό, καθώς και την κα Μπακατσάκη για την πολύτιμη υποστήριξή της. Επίσης, ευχαριστώ την οικογένειά μου για την στήριξή τους όλα αυτά τα χρόνια των σπουδών μου και όχι μόνο. Τέλος, θέλω να πω ένα ευχαριστώ στους συμφοιτητές-φίλους μου, για τα ωραιότερα φοιτητικά χρόνια που μου χάρισαν.

Στη μητέρα μου...

Περιεχόμενα

Περίληψη	iii
Ευχαριστίες	iv
1 Εισαγωγή	1
1.1 Στόχος Παρούσας Έρευνας	1
1.2 Δομή της Διπλωματικής Εργασίας	2
2 Σύνολα Δεδομένων	4
2.1 Δεδομένα	4
2.2 Ιστορική Αναδρομή	6
2.3 Μεγάλα Δεδομένα	8
2.4 Ανοιχτά Δεδομένα	14
2.5 Ανοιχτά Δημόσια Δεδομένα	16
2.6 Σχέσεις Αλληλεπίδρασης μεταξύ Μεγάλων Δεδομένων και Ανοιχτών Δεδομένων	19
3 Συστήματα Υποστήριξης Αποφάσεων	21
3.1 Η έννοια της πληροφορίας	21
3.2 Επιχειρηματική Ευφυΐα	24
3.2.1 Ορισμός και Τυπική Εφαρμογή της Επιχειρηματικής Ευφυΐας . .	24
3.2.2 Η αρχιτεκτονική ενός συστήματος Επιχειρηματικής Ευφυΐας . .	26
3.3 Βασικές Τεχνικές Εξόρυξης Πληροφοριών και Γνώσης	30

3.3.1	Κατηγοριοποίηση	30
3.3.2	Παλινδρόμηση	31
3.3.3	Ανάλυση Χρονοσειρών	31
3.3.4	Συσταδοποίηση	31
3.3.5	Αλγόριθμος K-means	32
3.4	Μειονεκτήματα - Πλεονεκτήματα Επιχειρηματικής Ευφυΐας	34
3.4.1	Μειονεκτήματα και Περιορισμοί	34
3.4.2	Πλεονεκτήματα και Οφέλη	35
4	Οπτικοποίηση Δεδομένων	37
4.1	Ταξινόμηση Μεθόδων Απεικόνισης Δεδομένων	38
4.2	Τεχνικές Οπτικοποίησης	40
4.3	Λογισμικό Tableau	47
4.3.1	Η εταιρεία Tableau Software Inc	48
4.3.2	Προϊόντα της εταιρείας	48
4.3.3	Βασικές Πληροφορίες Διεπαφής Χρήστη	50
5	Μελέτη Περίπτωσης Ασφαλιστικής Εταιρείας	54
5.1	Βάση Δεδομένων της Ασφαλιστικής Εταιρείας	55
5.2	Επεξεργασία και Προετοιμασία Βάσης Δεδομένων	58
5.3	Εξόρυξη Πληροφορίας Μέσω του Λογισμικού Tableau	59
6	Συμπεράσματα	73
	Βιβλιογραφία	75

Κεφάλαιο 1

Εισαγωγή

Η διαχείριση και αξιοποίηση δεδομένων, τα οποία παράγονται ολοένα και με υψηλότερους ρυθμούς, αποτελεί σοβαρή πρόκληση και αναγκαία αποστολή των σύγχρονων επιχειρήσεων. Αρκετές έρευνες έχουν πραγματοποιηθεί με σκοπό την ανάπτυξη μεθόδων επεξεργασίας και κατηγοριοποίησης των δεδομένων και βελτίωσης των συστημάτων της Επιχειρηματικής Ευφυΐας. Ο κυρίαρχος λόγος, για τον οποίο πυροδοτήθηκαν σχετικές μελέτες, σχετίζεται άμεσα με το γεγονός, ότι η ποσότητα των δεδομένων πολλαπλασιάζεται ραγδαία, η δυνατότητα εξόρυξης πληροφοριών αυξάνεται αναλογικά, παράλληλα όμως, οι απαιτήσεις για διαχείριση του υπέρογκου συνόλου δεδομένων είναι εξίσου υπαρκτές. Με βασικό γνώμονα τις ανάγκες που αναφέρθηκαν, αναπτύσσονται πλέον σύγχρονα συστήματα και τεχνολογίες που μπορούν να αξιοποιήσουν και να επεξεργαστούν ταχύτατα τεράστιες ποσότητες δεδομένων και να διεξάγουν συμπεράσματα, είτε υπό τη μορφή στατιστικών πινάκων, είτε υπό τη μορφή οπτικοποιήσεων.

1.1 Στόχος Παρούσας Έρευνας

Βασικός στόχος της διατριβής αυτής είναι, η ανάλυση των όρων Δεδομένα και Επιχειρηματική Ευφυΐα, καθώς αποτελούν θεμελιώδη στοιχεία της έρευνας και η αξιοποίηση ενός μεγάλου συνόλου δεδομένων, που αφορά πελάτες ασφαλιστικής εταιρείας, υπό τη

μορφή βάσης δεδομένων, μέσω της τεχνολογικής πλατφόρμας «Tableau». Αξιοποιώντας εμπράκτα το θεωρητικό υπόβαθρο της διατριβής (μέθοδος συσταδοποίησης δεδομένων (K-means), στατιστικοί πίνακες, οπτικοποιήσεις, κτλ), εξάγονται αποτελέσματα που δείχνουν τη σχέση της εταιρείας με το καταναλωτικό της κοινό.

1.2 Δομή της Διπλωματικής Εργασίας

Η διπλωματική εργασία εκπονείται σε έξι κεφάλαια. Αμέσως μετά το εισαγωγικό κεφάλαιο οδηγούμαστε στο δεύτερο κεφάλαιο, το οποίο αφορά συνολικά την έννοια του όρου δεδομένα και τους διαφορετικούς τύπους δεδομένων που υπάρχουν στη σύγχρονη εποχή. Περιγράφονται αναλυτικά, ο τρόπος παραγωγής και διάδοσης των δεδομένων, οι σύγχρονοι όροι των Μεγάλων Δεδομένων (Big Data), των Ανοιχτών και Δημόσιων δεδομένων και επισημαίνονται οι βασικές διαφορές που εμφανίζονται μεταξύ τους. Σκοπός του κεφαλαίου, είναι να δείξει με σαφήνεια την μεγάλη αξία και χρησιμότητα των πολυποίκιλων δεδομένων αλλά και να τονίσει την ανάγκη για εύστοχη διαχείριση του διαρκώς αυξανόμενου όγκου τους.

Το τρίτο κεφάλαιο, επικεντρώνεται στην Επιχειρηματική Ευφυΐα και τα συστήματα που αξιοποιεί, με κύριο στόχο την διαχείριση του τεράστιου όγκου δεδομένων που αναφέρθηκαν στο προηγούμενο κεφάλαιο. Πιο συγκεκριμένα, αναλύεται η αρχιτεκτονική (πυραμίδας) ενός συστήματος Επιχειρηματικής Ευφυΐας, περιγράφονται οι βασικές τεχνικές που χρησιμοποιούνται και αναδυνάμυνται τα κυριότερα μειονεκτήματα και πλεονεκτήματα των εν λόγω συστημάτων.

Τα περισσότερα σύγχρονα προγράμματα επεξεργασίας δεδομένων και εξόρυξης πληροφορίας, όπως το πρόγραμμα Tableau που χρησιμοποιείται στην παρούσα εργασία, εξάγουν τα αποτελέσματά τους υπό τη μορφή οπτικοποιήσεων. Για τον λόγο αυτό, στο τέταρτο κεφάλαιο περιγράφονται οι βασικές τεχνικές οπτικοποίησης των δεδομένων και

δίνονται οι απαραίτητες πληροφορίες σχετικά με το περιβάλλον διεπαφής του χρήστη στο πρόγραμμα του Tableau.

Στο πέμπτο κεφάλαιο, γίνεται πρακτική χρήση του προγράμματος «Tableau» μέσω του οποίου αξιοποιείται μεγάλη βάση δεδομένων πελατών ασφαλιστικής εταιρείας. Τα αποτελέσματα απαντούν περιγραφικά σε κρίσιμα ερωτήματα και επιδεικνύεται εμπεριστωμένα η σχέση της ασφαλιστικής εταιρείας με τους πελάτες της.

Στο έκτο και τελευταίο κεφάλαιο, αναφέρονται τα συμπεράσματα της διατριβής και προτίνονται ιδέες για περαιτέρω μελλοντική έρευνα.

Κεφάλαιο 2

Σύνολα Δεδομένων

Με την πάροδο των χρόνων, επιχειρηματίες, οργανισμοί ακόμη και κυβερνήσεις, ξεκίνησαν την συλλογή και επεξεργασία δεδομένων. Τα τελευταία χρόνια, ο όγκος των δεδομένων που συλλέγονται έχει αυξηθεί αλματωδώς, περισσότερο από οποιαδήποτε άλλη στιγμή στην ιστορία. Με τη ραγδαία τεχνολογική εξέλιξη, την πρόοδο των δικτυακών δυνατοτήτων και την ολοένα και μεγαλύτερη δυνατότητα πρόσβασης στο διαδίκτυο, τα δεδομένα αξιοποιούνται με κατάλληλες μεθόδους και παρέχουν κρίσιμες πληροφορίες. Στο παρόν κεφάλαιο θα αναλυθούν οι όροι των Δεδομένων, των Μεγάλων Δεδομένων και Ανοιχτών Δεδομένων, μιας και αποτελούν βασικές έννοιες σε θέματα λήψης αποφάσεων και εξόρυξης πληροφορίας, όπως επίσης θα τονιστούν οι ιδιαιτερότητες κάθε κατηγορίας δεδομένων αλλά και οι σχέσεις μεταξύ τους.

2.1 Δεδομένα

Τα δεδομένα αποτελούνται από ένα ακατέργαστο σύνολο στοιχείων που μπορεί να έχουν τη μορφή γραμμάτων ή λέξεων, αριθμών, συμβόλων, φωτογραφιών, βίντεο, διαγραμμάτων ακόμη και συνδυασμούς των προηγούμενων, εφόσον αυτό εξυπηρετεί τη διαδικασία καταγραφής τους, χωρίς να είναι απαραίτητα ξεκάθαρος ο σκοπός για τον οποίο αυτά συλλέγονται. Το σύνολο των καταχωρήσεων αυτών αποθηκεύεται ηλεκτρονικά μέσα σε ένα

σύστημα ηλεκτρονικού υπολογιστή, υπό τη μορφή ορισμένων bits. Πραγματοποιώντας τις απαραίτητες ενέργειες επεξεργασίας των ακατέργαστων δεδομένων και τεχνικές κωδικοποίησής τους, τα δεδομένα παίρνουν συνήθως τη μορφή στήλων και σειρών και κάθε θέση αποκτά μια ιδιότητα, μέσα στις οποίες καταχωρούνται οργανωμένα οι τιμές των δεδομένων. Η «τροφοδοσία» ενός προγράμματος στηρίζεται ουσιαστικά στην παροχή ηλεκτρονικών δεδομένων, με κυρίαρχο στόχο την παραγωγή χρήσιμων πληροφοριών (Stalidis et al., 2015).

Στη σύγχρονη οικονομία, το αυξανόμενο ενδιαφέρον των επιχειρήσεων για εύρεση, συλλογή και επεξεργασία δεδομένων, συνδέεται άρρηκτα με τους ραγδαίους ρυθμούς τεχνολογικής ανάπτυξης και κυρίως με τις σύγχρονες πλατφόρμες επεξεργασίας δεδομένων -πολλαπλών μορφών και δομών- που δεν έχουν περιορισμούς όπως τα παλαιότερα μοντέλα, που δυστυχώς παρουσίαζαν. Τα σύγχρονα αυτά εργαλεία έχουν την δυνατότητα να επεξεργαστούν τεράστιους όγκους δεδομένων και να πραγματοποιήσουν πολύπλοκους υπολογισμούς ταχύτητα, φέρνοντας μαζί τους την εποχή των Μεγάλων Δεδομένων. Σε έρευνα που διεξάγει η Διεθνής Εταιρεία Δεδομένων (IDC - International Data Corporation), προσδιορίζονται τρεις βασικές τοποθεσίες εξόρυξης δεδομένων, οι οποίες με τη σειρά τους συγκροτούν τη λεγόμενη *Σφαίρα των Δεδομένων* (Datasphere) (Reinsel et al., 2018).

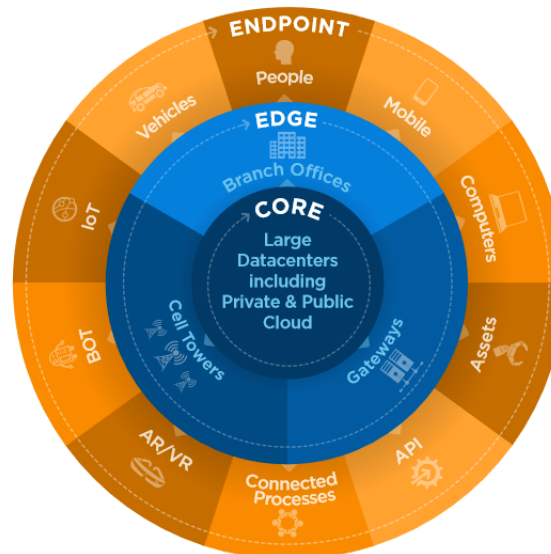
Η πρώτη τοποθεσία, ονομάζεται «**Core**» και αποτελείται από καθορισμένα κέντρα υπολογιστών, εντός της επιχείρησης, τα οποία είναι αρμόδια για την συλλογή και αποθήκευση δεδομένων. Στην ίδια τοποθεσία ανήκουν επίσης τα δεδομένα που συλλέγονται μέσω του cloud, μια παροχή υπολογιστικών υπηρεσιών δηλαδή, που επιτρέπει την αποθήκευση και διαχείριση βάσεων δεδομένων μέσω του διαδικτύου, «(ο.π.)».

Η δεύτερη τοποθεσία, ονομάζεται «**Edge**» και αφορά τα δεδομένα που παράγονται εκτός των καθορισμένων κέντρων στο εσωτερικό μιας επιχείρησης. Αναφορικά, τέτοιου είδους τοποθεσίες μπορούν να θεωρηθούν τα μικρότερα κέντρα δεδομένων, εκτός των

εγκαταστάσεων ενός οργανισμού, πύργοι κινητής τηλεφωνίας και δωμάτια διακομιστών, γνωστά ως server rooms (οπ.π.).

Τέλος, η τρίτη τοποθεσία ονομάζεται «**Endpoint**» και αφορά όλες τις συσκευές που υπάρχουν στο τέλος ενός δικτύου. Βασικοί χρήστες και δέκτες στην κατηγορία αυτή είναι οι πολίτες, μέσω των οποίων το μεγάλο σύνολο δεδομένων διαρκώς αυξάνεται και παρέχει άμεσα ή έμμεσα βελτίωση σε υπηρεσίες εξατομίκευσης, αναβαθμίσεις λογισμικών και εξυπηρέτησης πελατών, (οπ.π.).

Figure 2
Data propagation from endpoints to core and back



Source: IDC's Data Age 2025 study, sponsored by Seagate

Σχήμα 2.1: .
Τοποθεσίες Προέλευσης Δεδομένων (Reinsel et al., 2018)

2.2 Ιστορική Αναδρομή

Προτού προβούμε σε μια εκτενή ανάλυση της έννοιας των Μεγάλων Δεδομένων θα ήταν χρήσιμο να γίνει μια ιστορική αναδρομή, μέσω της οποίας θα αναδειχθούν η εξελικτική πορεία, τα αίτια και οι αφορμές που οδήγησαν στην συσσώρευση του τεράστιου όγκου

δεδομένων και στον τρόπο με τον οποίο αυτά επηρέασαν τη σύγχρονη επιχειρηματικότητα.

Στα τέλη της δεκαετίας 1980, τα συστήματα υποστήριξης αποφάσεων και αποθήκευσης δεδομένων σημείωσαν σημαντική ανάπτυξη. Οι δυνατότητες, που παρουσιάστηκαν από την πρόοδο αυτή, επεκτείνονται κυρίως στην ικανότητα των εταιρειών να πραγματοποιούν εκτενείς μετρήσεις, να παράγουν αναλύσεις προγνωστικού χαρακτήρα ακόμη και να καθορίζουν κανονικές τάσεις. Μέσα σ' αυτό το περιβάλλον αναδύθηκαν καινούριες λύσεις σε σύγχρονα προβλήματα, σχηματίστηκαν νέες εταιρείες που εξειδικεύτηκαν στην διαχείριση συστημάτων υποστήριξης αποφάσεων και η δημιουργία μιας νέας βιομηχανίας βασισμένη στη διαχείριση δεδομένων ήταν πλέον γεγονός (Krishnan, 2013, σ. 3-14).

Μερικά χρόνια αργότερα, το 1995 γίνονται οι πρώτες απόπειρες εμπορευματοποίησης του Διαδικτύου, με την δημιουργία των "dot-com" εταιρειών. Εταιρείες δηλαδή, που βασίζουν κυρίως την επιχειρηματικότητά τους και την αλληλεπίδρασή τους με τον πελάτη μέσω μιας ιστοσελίδας. Ο όγκος των δεδομένων σημειώνει μεγάλη αύξηση αρχίζοντας μάλιστα παράλληλα να εξατομικεύεται.

Το διαδικτυακό εμπόριο σε διάστημα λίγων ετών σημείωσε μεγάλη πρόοδο και διαμόρφωσε ένα νέο τοπίο για τους οργανισμούς, με σχηματισμό νέων επιχειρηματικών μοντέλων, που συναντάμε μέχρι και σήμερα. Συγκεκριμένα (ό.π.):

- **B2C** - Business to Consumer: Άμεση επαφή της Εταιρείας με τον ιδιώτη Καταναλωτή για αγοραπωλησίες, μέσω μίας ιστοσελίδας.
- **C2C** - Consumer to Consumer: Άμεση διαπραγμάτευση Καταναλωτή με Καταναλωτή. (eBay, Airbnb)
- **B2B2C** - Business to Business to Consumer: Έμμεση διαπραγμάτευση Εταιρείας με ιδιώτη Καταναλωτή, κατά την οποία μία μεσάζουσα εταιρεία αναλαμβάνει την προώθηση της υπηρεσίας/προϊόντων στο διαδίκτυο. (Google Play, App Store,

Amazon)

Παράλληλα, η δυνατότητα άσκησης κριτικής (reviews), η ανατροφοδότηση (feedback) μέσω σχολίων και εντυπώσεων για την συνολική εμπειρία του προϊόντος/υπηρεσίας και η εμφάνιση του μάρκετινγκ απο στόμα σε στόμα (WOM: word-of-mouth marketing) πολλαπλασίασαν ακόμη περισσότερο τον όγκο και το είδος των δεδομένων, αυξάνοντας όμως και τα επίπεδα πολυπλοκότητας, που αφορούν τη διαχείριση και επεξεργασία των δεδομένων αυτών.

Το έτος 2000 και η δεκαετία που ακολούθησε αποτέλεσαν καθοριστική περίοδο για την ιστορία των δεδομένων και του πρωταγωνιστικού ρόλου που κατέκτησαν στην σύγχρονη οικονομία. Με την εμφάνιση των κοινωνικών μέσων δικτύωσης (Facebook, Tumblr, Twitter, κτλ), των ισχυρών μηχανών αναζήτησης (Google, Yahoo), των εφαρμογών εξατομίκευσης μουσικής (Youtube), όπως επίσης και το γεγονός ότι τα περισσότερα κινητά τηλέφωνα, ηλεκτρονικοί υπολογιστές και Tablet μπορούν να συνδέονται στο διαδίκτυο μέσω WiFi ή 3G δικτύων ενίσχυσαν εκρηκτικά την συσσώρευση τεράστιων όγκων δεδομένων και κατέστησαν αναγκαία τη δημιουργία τρόπων και μεθόδων ισχυρών μετρήσεων και διεργασιών, ώστε να αξιοποιηθεί η πληροφορία και η γνώση που κρύβεται ακατέργαστη (όπ.π.).

2.3 Μεγάλα Δεδομένα

Με τον όρο Μεγάλα Δεδομένα ορίζονται οι μεγάλοι όγκοι δεδομένων που διατίθενται σε ποικίλους βαθμούς πολυπλοκότητας, οι οποίοι παρουσιάζουν πολύμορφες ασάφειες και η διαδικασία παραγωγής τους πραγματοποιείται σε ανόμιες ταχύτητες (Krishnan, 2013, σ. 3-14). Η φύση και το εύρος αυτών των δεδομένων, καθιστούν τις παραδοσιακές μεθόδους ανάλυσης και τις τεχνολογίες που τις χαρακτηρίζουν ανεπαρκείς, ως προς τη

δυνατότητά τους να διαχειριστούν τα Μεγάλα Δεδομένα, τόσο από άποψη ποσότητας όσο και από άποψη ταχύτητας επεξεργασίας.

Η αναγνώριση των Μεγάλων Δεδομένων και της αδιαμφισβήτητης σημασίας τους για οποιαδήποτε διαδικασία λήψης αποφάσεων έγινε αρχικά, σε σχετικό άρθρο, από την McKinsey & Co., όμως η θεμελίωση του όρου πραγματοποιήθηκε από τον γνωστό αναλυτή Doug Laney που βασίστηκε για την ανάλυση του όρου σε πέντε ("V") του αγγλικού αλφαβήτου(οπ.π.).

Πιο αναλυτικά:

- **Volume - Όγκος:**

Η παραγωγή δεδομένων είναι μια διαδικασία που δεν διακόπτεται, ο όγκος των Μεγάλων Δεδομένων σημειώνει διαρκώς αυξητική τάση και η αξιοποίηση της τεχνολογίας των δεδομένων από οργανισμούς είναι πλέον πολύ συχνό φαινόμενο. Τα Μεγάλα Δεδομένα συλλέγονται και ύστερα αποθηκεύονται σε τεράστιες ψηφιακές αποθήκες δεδομένων και αξιοποιούνται κατάλληλα για να παρέχουν σε επιχειρήσεις και οργανισμούς μεγαλύτερη ευελιξία στην αγορά, βελτιωμένη σχέση με τον πελάτη, ανάπτυξη νέων επιχειρηματικών μοντέλων και σωστή διαχείριση ρίσκου.

Βασικές πηγές παραγωγής δεδομένων θεωρούνται:

- Έξυπνες Συσκευές
- Επιχειρηματικές Συναλλαγές
- Μέσα Κοινωνικής Δικτύωσης
- Δίκτυα Αισθητήρων
- Σαρωτές και συσκευές ακτινών X-ray
- Οπτικοακουστικό Υλικό (Video)
- Βιομηχανικό Υλικό

Παράλληλα κάθε οργανισμός μπορεί να διασφαλίσει τη δική του παραγωγική διαδικασία δεδομένων, η οποία συνήθως περιλαμβάνει:

- Νομικά Θέματα
- Marketing
- Χρηματοοικονομικά θέματα
- Προμήθειες
- Πωλήσεις ή Αγορές
- Ενέργειες που αφορούν το ανθρώπινο δυναμικό

- **Variety** - Ποικιλία:

Οι τεράστιοι όγκοι δεδομένων χαρακτηρίζονται από μεγάλη ποικιλομορφία σε ότι αφορά το είδος και το τύπο τους. Το κυριότερο μέρος των Μεγάλων Δεδομένων αποτελείται από στοιχεία τα οποία δεν έχουν συγκεκριμένη δομή. Είναι ήχοι, βίντεο, e-mails, κείμενα, δεδομένα μετοχών ακόμη και χρηματοοικονομικές συναλλαγές. Αυτός ο καινούριος όγκος δεδομένων απαιτούσε νέες τεχνολογίες και αλγόριθμους που θα μπορούσαν να αναλύσουν και επεξεργαστούν την αυθαίρετη δομή τους και τις πολύμορφες πηγές από όπου γίνεται η εξόρυξή τους. Τα Μεγάλα Δεδομένα λοιπόν θα μπορούσαν να χωριστούν σε δύο κατηγορίες ανάλογα τον τύπο τους, (Nugent et al., 2013).

Στην πρώτη κατηγορία ανήκουν τα δεδομένα με συγκεκριμένη δομή. Ως *δομημένα* ονομάζονται τα δεδομένα που διαθέτουν καθορισμένο μήκος και μορφή. Χαρακτηριστικά παραδείγματα στην περίπτωση αυτή αποτελούν καταχωρήσεις όπως ονόματα, ηλικίες, ημερομηνίες, μήνες και συμβολοσειρές, δηλαδή ακολουθίες λέξεων και αριθμών που μαζί αποκτούν συγκεκριμένο νόημα. Η αποθήκευση των δεδομένων γίνεται συνήθως σε μια βάση δεδομένων, είτε σε κάποιο υπολογιστικό φύλλο και

αποτελούν την πιο οικεία μορφή δεδομένων στο μέσο χρήστη. Οι καταχωρήσεις μπορούν να πραγματοποιηθούν είτε από ανθρώπινη παρέμβαση είτε από οποιοδήποτε μηχανικό μέσο. Τα συγκεκριμένης δομής δεδομένα αποτελούν μόλις το 20% του συνόλου των Μεγάλων δεδομένων και πολλοί οργανισμοί χρησιμοποιούν μέχρι και σήμερα παραδοσιακούς τρόπους συλλογής τους.

Η δεύτερη κατηγορία, αποτελείται από το υπόλοιπο 80% του συνόλου των Μεγάλων Δεδομένων και αφορά δεδομένα τα οποία δεν έχουν συγκεκριμένη δομή. Τα *μη-δομημένα* δεδομένα δημιουργούνται και καταγράφονται σε όλες τις πτυχές της σύγχρονης ζωής. Η καθημερινότητα των ανθρώπων κατακλύζεται από αδόμητα δεδομένα και η επιχειρηματική δράση μέσα στους οργανισμούς δεν αποτελεί εξαίρεση. Η καταγραφή των αδόμητων δεδομένων επιτυγχάνεται μέσω του ανθρώπινου παράγοντα, είτε μέσω κάποιου μηχανικού μέσου (οπ.π.).

Μερικά παραδείγματα δεδομένων μη συγκεκριμένης δομής:

- Εικόνες του υπεδάφους και των σεισμικών πλακών
- Ζωντανή αναμετάδοση από δορυφόρους για έγκυρη και έγκαιρη πρόγνωση του καιρού
- Καταγραφή οπτικοακουστικού υλικού από συσκευές παρακολούθησης, κυρίως με σκοπό την επιτήρηση και την ρύθμιση κυκλοφορίας
- Κείμενα-φωτογραφίες-ήχοι από επίσκεψη σε σελίδες μέσων κοινωνικής δικτύωσης
- Δεδομένα από διάφορα site ή blogs
- Προφορικός Λόγος

Αναφέρεται επίσης, η περίπτωση των *ημι-δομημένων* δεδομένων. Δεδομένα δηλαδή, τα οποία εμπίπτουν μεταξύ των δομημένων και μη δομημένων τύπων. Τα

δεδομένα αυτής της κατηγορίας δεν υπόκεινται απαραίτητα σε μια συγκεκριμένη μορφή, μπορούν όμως να βοηθήσουν τους αναλυτές και τις οργανώσεις να προβούν σε περαιτέρω επεξεργασία των δεδομένων και να κατανοήσουν σε μεγαλύτερο βάθος πολυσύνθετα συμβάντα, (ο.π.).

- **Value - Αξία:**

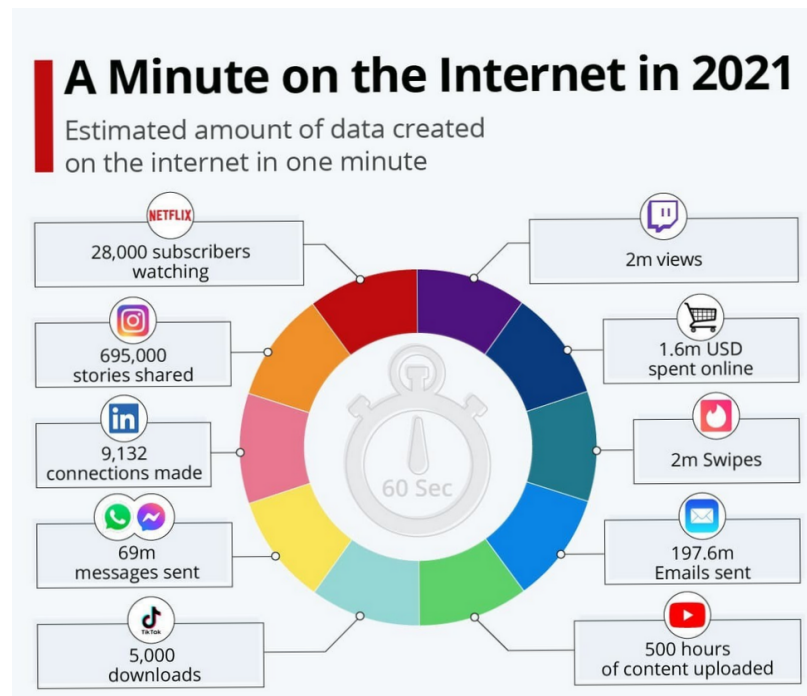
Τα δεδομένα αντιμετωπίζονται ως προϊόν και η αξία τους καθορίζει σε μεγάλο βαθμό τις ευκαιρίες και τις εναλλακτικές δράσεις που μια οργάνωση καλείται να διαλέξει. Η ποιότητα του αποτελέσματος είναι άρρηκτα συνδεδεμένη και ανάλογη της αξίας, που αντλήθηκε από τα δεδομένα που χρησιμοποιήθηκαν. Σαφώς, η ικανότητα της ομάδας των αναλυτών και οι τεχνολογίες που χρησιμοποιούνται για την διαδικασία εξόρυξης πληροφοριών παίζουν σημαντικό ρόλο σε κάθε βήμα της διαδικασίας, όμως πρέπει να τονιστεί πως η σημαντικότερη προϋπόθεση για τη διεξαγωγή έγκυρων πληροφοριών που θα παρέχουν καινοτόμες ιδέες και αποφάσεις, βασίζεται στην αξία των δεδομένων και την μοναδικότητα που χαρακτηρίζει τον τρόπο με τον οποίο αντλήθηκε η αξία αυτή. Η συλλογή και χρήση των δεδομένων από έναν οργανισμό θα πρέπει να γίνεται με ιδιαίτερη προσοχή και σύνεση, αφού τα συμπεράσματα των αναλύσεων ίσως θέσουν τα μελλοντικά σχέδια, στόχους και στρατηγικές προτεραιότητες των επιχειρήσεων.

- **Velocity - Ταχύτητα:**

Ως ταχύτητα των Μεγάλων Δεδομένων εννοείται ο ρυθμός με τον οποίο τα δεδομένα παράγονται, συλλέγονται και αποθηκεύονται στις κατάλληλες αποθήκες δεδομένων. Με γνώμονα την ταχύτητα των δεδομένων δημιουργούνται δύο βασικές κατηγορίες. Η πρώτη κατηγορία, αφορά δεδομένα τα οποία παράγονται σε πραγματικό χρόνο, η ταχύτητα παραγωγής τους είναι υψηλή και συνήθως σχετίζεται με κείμενα, σχόλια και αντιδράσεις χρηστών σε πλατφόρμες κοινωνικής δικτύωσης ή αντίστοιχους ιστότοπους συχνής επισκεψιμότητας και αυξημένης δραστηριότητας.

Η δεύτερη κατηγορία οργανώνει τα δεδομένα σε παρτίδες, δεν υπάρχει δηλαδή μια αδιάκοπη ροή δεδομένων, τα οποία αποστέλονται την στιγμή που καταγράφονται. Αντιθέτως, η διαδικασία έχει αρχικό και τελικό στάδιο και το προϊόν που προκύπτει αποτελείται από ένα σύνολο δεδομένων και όχι από επιμέρους «μονάδες» δεδομένων, (οπ.π.).

Χαρακτηριστικό παράδειγμα υψηλών ταχυτήτων παραγωγής δεδομένων φαίνεται στο Σχήμα[2.2], στο οποίο αναγράφεται η εκτιμώμενη ποσότητα δεδομένων που δημιουργείται με το πέρας ενός λεπτού της ώρας.



Σχήμα 2.2:

Εκτιμώμενη ποσότητα δεδομένων που παράγεται στο διαδίκτυο, για κάθε λεπτό της ώρας που περνά. (Claire Jenik, 2021)

- **Veracity** - Εγκυρότητα:

Η έννοια της εγκυρότητας στο χώρο των Μεγάλων Δεδομένων προσδιορίζεται από συγκεκριμένους παράγοντες που στόχο έχουν την ακρίβεια, την ποιότητα αλλά και την αξιοπιστία της πηγής των δεδομένων, από την οποία προέρχονται. Βασικός στόχος μιας ανάλυσης με γνώμονα τον όρο «Veracity», είναι η εύστοχη πρόβλεψη της επιχειρηματικής αξίας την οποία κρύβει οποιαδήποτε επιχειρηματική δράση που πρόκειται να πραγματοποιηθεί. Παράλληλα, όμως γίνεται αυστηρός έλεγχος για τον τύπο των δεδομένων, τον τρόπο που αυτά επεξεργάζονται και τελικά αν η ανάλυση των Μεγάλων Δεδομένων οδηγεί σε κάποιο πρακτικό νόημα. Σε αυτό το σημείο, πρέπει να σημειωθεί η ανάγκη των οργανισμών να έχουν στη διάθεσή τους τις απαραίτητες τεχνολογίες και την απαραίτητη τεχνογνωσία να αποθηκεύουν, επεξεργάζονται και να αξιοποιούν όλων των τυπών και μορφών δεδομένα. Χαρακτηριστικό παράδειγμα μίας τέτοιας διαδικασίας αποτελεί η πρόγνωση σοβαρών καταγίδων και η εξέλιξη της συνολικής τους πορείας. Συγκεκριμένα, συσχετίζονται άμεσα ροές δεδομένων του ιστότοπου Twitter με δεδομένα τηλεμετρίας από έναν μετεωρολογικό δορυφόρο και με αυτό τον τρόπο γίνεται γνωστή τόσο η αφετηρία και η εξέλιξη του καιρικού φαινομένου όσο και ο τρόπος με τον οποίο επηρεάζεται ο τοπικός πληθυσμός καθ' όλη τη διάρκεια εξέλιξής του, (ο.π.).

2.4 Ανοιχτά Δεδομένα

Ο όρος Ανοιχτά Δεδομένα αναφέρεται σε ένα μεγάλο σύνολο δεδομένων που αποθηκεύονται και χρησιμοποιούνται ελεύθερα, επαναχρησιμοποιούνται και διανέμονται από οποιονδήποτε χρήστη εξίσου ελεύθερα, με βασική προϋπόθεση την αναφορά του δημιουργού των δημοσιοποιημένων δεδομένων που χρησιμοποιήθηκαν.

Η έννοια των Ανοιχτών Δεδομένων και ανοιχτού περιεχομένου που βρίσκεται στον ιστότοπο του "Open Definition", συνοψίζεται στον εξής ορισμό (Open Knowledge Foundation, 2022):

«Ανοιχτό σημαίνει πως ο καθένας μπορεί να έχει ελεύθερη πρόσβαση, να κάνει χρήση, να επεξεργάζεται και να μοιράζεται για οποιοδήποτε σκοπό με την τήρηση των προϋποθέσεων που διαφυλάσσουν την προέλευση και τη διαφάνεια.»

Ακολούθως, παρουσιάζονται περαιτέρω πληροφορίες σε ό,τι αφορά τα Ανοιχτά Δεδομένα:

- Προσβασιμότητα και διαθεσιμότητα: Τα δεδομένα θα πρέπει να είναι αμετάβλητα και αθέμελα. Η λήψη τους ιδανικά, θα πρέπει να πραγματοποιείται μέσω διαδικτύου και το κόστος αναπαραγωγής τους θα πρέπει να είναι λογικό. Βασικό στοιχείο που πρέπει να τονισθεί είναι η κατάσταση στην οποία θα βρίσκονται τα Ανοιχτά Δεδομένα. Θα πρέπει δηλαδή, τα δεδομένα να είναι τέτοιας μορφής που θα επιτρέπεται η άμεση και πλήρης ανάγνωσή τους από υπολογιστικές μηχανές. Τέλος, οι όροι που θα διέπουν τα δεδομένα θα πρέπει να είναι τέτοιοι ώστε να επιτρέπουν την επαναχρησιμοποίηση και αναδιανομή τους, ακόμη και αν αυτά αξιοποιήθηκαν σε συνδυασμό με άλλα σύνολα δεδομένων.
- Πρόσβαση δίχως διακρίσεις: Η χρήση των Ανοιχτών Δεδομένων δεν μπορεί να περιορίζεται από συγκεκριμένα άτομα, ομάδες ή ειδικούς επιστημονικούς κύκλους. Οι αυστηρές προϋποθέσεις χρήσεως, υπό την ύπαρξη περιορισμών δεν γίνονται επιτρεπτές καθώς τα Ανοιχτά Δεδομένα ανθαρρύνουν την καθολική συμμετοχή, χωρίς διακρίσεις και φραγμούς.
- Είδη Ανοιχτών Δεδομένων: Τα είδη των Ανοιχτών Δεδομένων ποικίλουν και έχουν πολλαπλές εφαρμογές και χρήσεις και οδηγούν στη δημιουργία των παρακάτω κατηγοριών:

- Πολιτισμός: Πολιτιστικά έργα και αντικείμενα. Δεδομένα που συλλέγονται κυρίως απο αρχεία, μουσεία ή γκαλερί.
- Επιστήμη: Παράγονται ως μέρος επιστημονικής έρευνας και καλύπτουν ολόκληρο το φάσμα των επιστημών.
- Οικονομικά: Έσοδα - έξοδα κρατικών λογαριασμών, πληροφορίες για χρηματοπιστωτικές συναλλαγές (μετοχές, ομόλογα, κτλ).
- Στατιστικά: Απογραφές, κοινωνικοοικονομικοί δείκτες και διάφορα δεδομένα που παράγονται απο στατιστικές υπηρεσίες.
- Καιρός: Ποικίλες πηγές συλλογής Ανοιχτών Δεδομένων για την κατανόηση και πρόγνωση καιρικών φαινομένων, καθώς και του κλίματος.
- Περιβάλλον: Επίπεδα ρίπων, στάθμη των υδάτων, κτλ.

Η αποσαφήνιση του όρου των Ανοιχτών Δεδομένων έχει μεγάλη σημασία, διότι μέσω της ελευθερίας που διασφαλίζεται μπορεί και επιτυγχάνεται *διαλειτουργικότητα*. Με τον όρο διαλειτουργικότητα εννοείται η δυνατότητα πολλών διαφορετικών συστημάτων να συνδυάζονται ή να συγχωνεύονται μεταξύ τους. Αυτή η δυνατότητα διαμοιρασμού και αλληλοσύνδεσης συνιστωσών αποτελεί βασικό θεμέλιο για τη δόμηση και επεξεργασία πιο πολυσύνθετων και πολύπλοκων συστημάτων.

Τέλος, τα Ανοιχτά Δεδομένα, δεν θα πρέπει να περιλαμβάνουν προσωπικά δεδομένα, οποιασδήποτε μορφής. Απαιτείται, δηλαδή να τηρούνται όλοι οι κανόνες προστασίας προσωπικών δεδομένων και να μην μοιράζονται πληροφορίες και στοιχεία ατόμων.

2.5 Ανοιχτά Δημόσια Δεδομένα

Η σύγχρονη προσέγγιση για κοινή χρήση, ελεύθερη πρόσβαση, επεξεργασία και επανα-προώθηση δεδομένων, ενισχύει την ανάπτυξη ψηφιακής διαλειτουργικότητας, διευρύνει τα συνεργατικά δίκτυα και εδραιώνει ένα περιβάλλον αστείρευτης ανταλλαγής πληροφο-

ριών σε καθημερινή βάση. Μία νέα, σύγχρονη κινητήρια δύναμη που υπόσχεται να φέρει κοινωνικές, οικονομικές και πολιτικές αλλαγές οφείλεται στη συλλογή και επεξεργασία των *Ανοιχτών Δημόσιων Κυβερνητικών Δεδομένων*.

Πρόκειται για την ελεύθερη διάθεση των δεδομένων που η δημόσια διοίκηση παράγει, συλλέγει και διαχειρίζεται και τα οποία αφορούν την οικονομική και την κοινωνική ζωή των πολιτών. Στο σύγχρονο πολιτισμό, όπου η ελεύθερη διακίνηση και ανάλυση δεδομένων μέσω προσβάσιμων τεχνολογιών είναι γεγονός, το μονοπώλιο συλλογής δεδομένων, που ιστορικά διατηρούσε για μεγάλο χρονικό διάστημα η κυβέρνηση, υπονομεύεται και με σταθερούς ρυθμούς καταργείται. Μάλιστα οι πιέσεις για δημοσιοποίηση και ελεύθερη πρόσβαση στα δεδομένα που συγκεντρώνουν οι κυβερνήσεις είναι έντονες, με βασικούς πρωταγωνιστές, στις δράσεις αυτές, να είναι λόμπι από το χώρο Πληροφοριών Δημόσιου Τομέα (P.S.I - Public Sector Information), με στόχο την επαναχρησιμοποίηση των δεδομένων για εμπορικούς σκοπούς και ομάδες ακτιβιστών που υποστηρίζουν θερμά την ελευθερία της πληροφόρησης (F.O.I - Freedom of Information). Τα τελευταία 10 χρόνια, οι προσπάθειες για ανοιχτή διακυβέρνηση έχουν αποδώσει καρπούς και η πρόσβαση σε κυβερνητικά σύνολα δεδομένων είναι εντυπωσιακή. Η αξιοποίηση Ανοιχτών Δημόσιων Δεδομένων βρίσκεται σε πειραματικό στάδιο, μεγάλο όμως ενδιαφέρον παρουσιάζεται στα αποτελέσματα ερεύνas που ανέδειξαν τις ομάδες χρηστών με μεγαλύτερο ενδιαφέρον για αξιοποίηση των κυβερνητικών δεδομένων. Συγκεκριμένα, σε δημογραφική έρευνα που πραγματοποιήθηκε στις αρχές του 2010, η συντριπτική πλειοψηφία χρηστών Ανοιχτών Δημόσιων Δεδομένων ήταν άνδρες, με την αξιοσημείωτη αναλογία 6 προς 1 έναντι των γυναικών. Τα περισσότερα άτομα άνηκαν στο τομέα των μικροεπιχειρήσεων και των ιδιωτικών ΜΜΕ, σε τοπικά και εθνικά ιδρύματα δημόσιου τομέα και σε ακαδημαϊκά ιδρύματα (Davies, 2010, σ. 1-47).

Η ανάλυση των χρηστών Ανοιχτών Δημόσιων Δεδομένων, με βάση τα κίνητρα τους, δημιουργεί 6 ξεχωριστές ομάδες με διαφορετικές επιδιώξεις και συμφέροντα (ο.π.):

- Κυβερνητική Εστίαση: Με πρωταρχικό στόχο την βαθύτερη κατανόηση της λειτουργίας και της αποστολής που φέρει μια κυβέρνηση ενισχύεται η λογοδοσία και η αποτελεσματικότητα της.
- Τεχνολογικές Καινοτομίες: Ανάπτυξη νέων πλατφόρμων και τεχνολογιών γύρω από την επιστήμη των δεδομένων.
- Ανταμοιβή: Διαχείριση των δεδομένων με απώτερο σκοπό την αναγνώριση ή την επίτευξη κέρδους.
- Ψηφιοποίηση: Τεχνολογικές βελτιώσεις και καινοτομίες για εξυγίανση και αποτελεσματικότερη διακυβέρνηση.
- Επίλυση Προβλήματος: Χρήση του συνόλου Ανοιχτών Δημόσιων Δεδομένων με σκοπό την ανάλυση προβλημάτων και σχηματισμού λύσεων.
- Δημόσιος Τομέας: Βελτίωση παροχής υπηρεσιών στο δημόσιο τομέα.

Η Ευρωπαϊκή Ένωση έχει δημιουργήσει μία απόθήκη Ανοιχτών Δεδομένων γνωστή ως «Ευρωπαϊκή Πύλη Δεδομένων», στην οποία συλλέγονται τα μεταδεδομένα των πληροφοριών του δημόσιου τομέα, που είναι διαθέσιμα σε πύλες δημόσιων δεδομένων στην επικράτεια των ευρωπαϊκών χωρών.

Από την χρήση των Ανοιχτών Δεδομένων γίνεται εξόρυξη πληροφοριών και γνώσης με πολλαπλά οφέλη στη βελτίωση της απόδοσης δημοσίων διοικήσεων, τη διεύρυνση του κράτους πρόνοιας και την ανάπτυξη νέων επιχειρηματικών μοντέλων, για κάθε χώρα ξεχωριστά. Παράλληλα, πραγματοποιείται βελτίωση της αποδοτικότητας των δημοσίων υπηρεσιών, με έλεγχο και περιορισμό των περιττών δαπανών.

Μεγαλύτερη απόδοση, εξοικονόμηση χρόνου, μείωση ενεργειακής κατανάλωσης α-κόμη και αποφυγή τροχαίων ατυχημάτων στους δρόμους είναι μερικά από τα πλεονεκτήματα που προκύπτουν από την ανάλυση Ανοιχτών Δεδομένων. Συγκεκριμένα, 7.000

ζωές σώζονται ετησίως επιταχύνοντας την παροχή ανάνηψης σε θύματα, εξοικονομούνται 629 εκατομμύρια ώρες περιττού χρόνου αναμονής σε δρόμους της Ευρωπαϊκής Ένωσης, 2.549 ώρες καταναλώνονται αποκλειστικά στην προσπάθεια εύρεσης χώρου στάθμευσης οχημάτων και σημειώνεται 16% μείωση της κατανάλωσης ηλεκτρικής ενέργειας (Europa, 2022).

2.6 Σχέσεις Αλληλεπίδρασης μεταξύ Μεγάλων Δεδομένων και Ανοιχτών Δεδομένων

Αδιαμφισβήτητα, τόσο το σύνολο των Μεγάλων Δεδομένων όσο και το σύνολο των Ανοιχτών Δεδομένων, αποτελούν φαινόμενο της σύγχρονης εποχής και η χρηστική τους αξία δεν έχει περάσει απαρατήρητη. Στην περίπτωση των Μεγάλων Δεδομένων διακρίνουμε τεράστιες ποσότητες δεδομένων, που τις περισσότερες φορές αμφισβητούνται τα όρια των ικανοτήτων του ανθρώπου να τις αναλύσει. Στην περίπτωση όμως των Ανοιχτών Δεδομένων, η πολυπλοκότητα και το μέγεθος του συνόλου δεν αποτελούν τα βασικά στοιχεία που το καθορίζουν. Τα Ανοιχτά δεδομένα δημιουργήθηκαν και κυκλοφόρησαν με πρωταρχικό στόχο να παρέχονται σκόπιμα για δημόσια χρήση. Η πρόσβαση στα Ανοιχτά Δεδομένα είναι ελεύθερη προς όλους τους ανθρώπους ανεξαιρέτως τις ομάδες ή τους οργανισμούς και οι επιπτώσεις τελικά που οι επιπτώσεις που προκύπτουν από την διαχείριση και επεξεργασία τους, δεν αφορούν μικρά και περιορισμένα σύνολα ή πλυθυσμούς, καθώς τα δεδομένα μοιράζονται στο ευρύ κοινό και αποτελούν δημόσιο αγαθό. Στη συνέχεια, παρατείνεται διάγραμμα τύπου **Venn**, Σχημα[2.3], που δείχνει τη σχέση μεταξύ Μεγάλων και Ανοιχτών Δεδομένων.

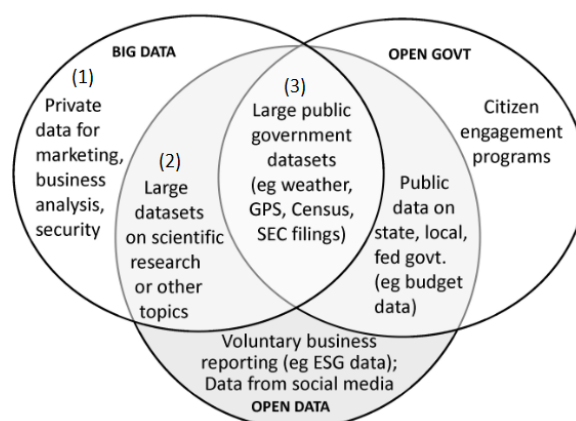
Το **χωρίο (1)**, αποτελείται από Μεγάλα Δεδομένα, τα οποία δεν είναι ανοιχτά. Είναι δεδομένα που συλλέγονται είτε από τις κυβερνήσεις για συγκεκριμένα άτομα ή ομάδες και αφορούν την εθνική ασφάλεια είτε από εμπόρους για βαθύτερη κατανόηση της συμπεριφοράς των πελατών τους. Τα δεδομένα αυτού του χωρίου, τι περισσότερες

φορές, αποτελούν πηγή δημόσιας ανησυχίας, καθώς καταπατούν αρκετές φορές βασικούς κανόνες ιδιωτικότητας και προστασίας προσωπικών δεδομένων (Gurin, 2015).

Το **χωρίο (2)**, αποτελείται από Μεγάλα Δεδομένα που είναι συγχρόνως ανοιχτά. Αφορά επιστημονικά δεδομένα, τα οποία είναι διαθέσιμα προς επεξεργασία από οποιαδήποτε ερευνητική ομάδα ή επιστήμονα. Πρόκειται για βάσεις δεδομένων απο επιστημονικές έρευνες και η διαδικασία ελεύθερης πρόσβασης και κοινής χρήσης επιταχύνει πολλές φορές την πρόοδο των ερευνών και την διεξαγωγή αποτελεσμάτων (ο.π.).

Τέλος, το **χωρίο (3)**, αποτελείται από μεγάλα σύνολα κυβερνητικών δεδομένων, ανοιχτού χαρακτήρα. Βασικός σκοπός του συνόλου δεδομένων του χωρίου αυτού, είναι η αναβάθμιση της ποιότητας ζωής των πολιτών μιας κοινωνίας και η δημιουργία ευκαιριών για νέες επιχειρηματικές δράσεις (ο.π.).

Η ανάλυση Μεγάλων Δεδομένων και η χρήση τους ως Ανοιχτά Δεδομένα χαράσσουν νέους κοινωνικοοικονομικούς ορίζοντες και σηματοδοτούν μία νέα εποχή προόδου και καινοτομίας. Θέτοντας τα δεδομένα στο επίκεντρο των διεργασιών και αξιοποιώντας συστήματα επιχειρηματικής ευφυΐας, μεθόδους ανάλυσης δεδομένων και οπτικοποίησης, οδηγούμαστε στην κατάκτηση ανεκτίμητης πληροφορίας και γνώσης.



Σχήμα 2.3: .

Διάγραμμα τύπου Venn σχέσης μεταξύ Μεγάλων και Ανοιχτών Δεδομένων (Gurin, 2015).

Κεφάλαιο 3

Συστήματα Υποστήριξης Αποφάσεων

Τα Μεγάλα Δεδομένα αποτελούνται από τεράστια σύνολα δομημένων και μη-δομημένων δεδομένων και μαζί σχηματίζουν ένα πολύτιμο σύνολο ακατέργαστων πληροφοριών, το οποίο χρήζει ανάλυσης και επεξεργασίας, με αξιοποίηση αλγορίθμων και προηγμένων μεθόδων ανάλυσης δεδομένων. Στο παρόν κεφάλαιο αναλύονται οι όροι της *Επιχειρηματικής Ευφυΐας* και των συστημάτων που αξιοποιεί, περιγράφονται σημαντικές τεχνικές Εξόρυξης Πληροφορίας και εντοπίζονται τα κύρια μειονεκτήματα και πλεονεκτήματα των συστημάτων υποστήριξης αποφάσεων που βασίζονται σε μεθόδους και πρακτικές της *Επιχειρηματικής Ευφυΐας*.

3.1 Η έννοια της πληροφορίας

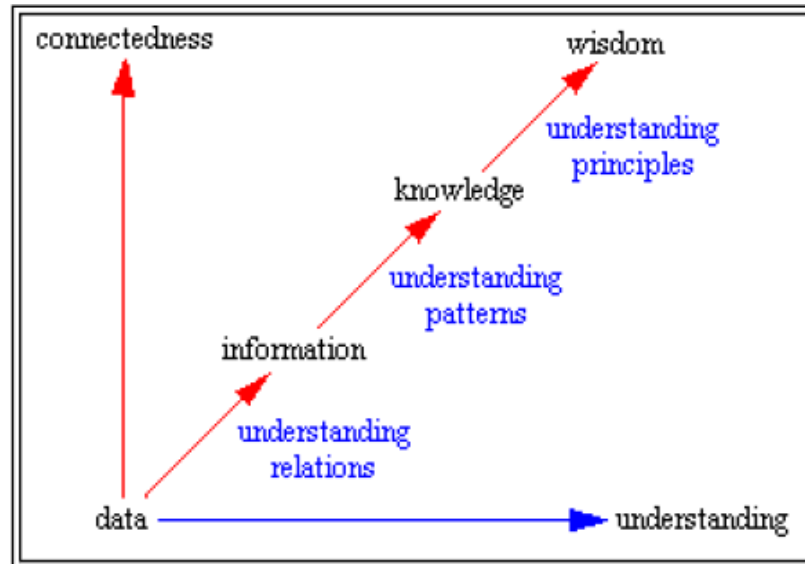
Είναι σαφές πως ο όρος των δεδομένων και ο όρος της πληροφορίας διαφέρουν μεταξύ τους. Αρκετές βέβαια είναι οι περιπτώσεις στις οποίες δεν αποσαφηνίζονται επαρκώς οι διαφορές των δύο εννοιών και παρά το γεγονός ότι υπάρχει έλλειψη ενός αποδεκτού συνόλου ορισμών, τα δεδομένα δεν μπορούν να θεωρούνται πληροφορία και οι πληροφορίες αυτούσιες, δεν αποτελούν γνώση.

Σε ένα πληροφοριακό σύστημα είναι δυνατόν να οριστούν τέσσερα (4) επίπεδα διάκρισης πληροφοριών, με την μορφή και το είδος της κάθε πληροφορίας να προσδιορίζεται από το επίπεδο στο οποίο ανήκει. Ειδικότερα, το πρώτο επίπεδο εμπεριέχει στοιχεία κωδικοποιημένα, που διευκολύνουν περισσότερο μια υπολογιστική μηχανή να «διαβάσει» και να επεξεργαστεί τα στοιχεία αυτά. Φτάνοντας στο τέταρτο επίπεδο, τα στοιχεία πληροφοριών που εμπεριέχονται είναι εγγύτερα στην ανθρώπινη λογική και χαρακτηρίζονται από υψηλή οργάνωση και ευφυΐα. Τα επίπεδα έχουν βασιστεί, κυρίως, στην άρρηκτη και αναγκαία σχέση μεταξύ των δεδομένων - πληροφορίας - γνώσης και του τελικού σταδίου που αφορά την κατάκτηση σοφίας, (Stalidis et al., 2015).

- **1ο Επίπεδο - Δεδομένα:** Ως πρώτη ύλη και σημείο εκκίνησης θεωρούνται τα δεδομένα. Η αποθήκευση τους σε έναν υπολογιστή γίνεται σε μορφή bit και η αναπαράσταση τους, μέσω λογισμικών, μπορεί να γίνεται κατανοήτη και διαχειρίσιμη από τον άνθρωπο.
- **2ο Επίπεδο - Πληροφορία:** Η δυνατότητα να αποδίδεται νόημα και σχεσιακή σύνδεση μεταξύ των δεδομένων, επί της ουσίας αποτελεί και την παραγωγική διαδικασία απόκτησης πληροφοριών.
- **3ο Επίπεδο - Γνώση:** Μια σωστή απάντηση σε ένα ερώτημα απαιτεί μία βαθιά αναλυτική ικανότητα, που περιλαμβάνει γύρω από ένα υψηλό επίπεδο κατανόησης. Η εύστοχη διαλογή πληροφοριών και η ικανότητα να χρησιμοποιηθούν αυτές με πρακτικό όφελος, για την επίλυση επιχειρηματικών προβλημάτων, ορίζουν την έννοια της γνώσης. Στην περίπτωση των υπολογιστών, οι περισσότερες εφαρμογές που χρησιμοποιούνται, βασίζονται στην επιτυχημένη λειτουργία τους σε κάποιο είδος αποθηκευμένης γνώσης (Bellinger et al., 2004)
- **4ο Επίπεδο - Σοφία:** «Σύνολο από γνώσεις και εμπειρία που συνδυάζεται με δυνατότητα κρίσης και μπορεί να εφαρμοστεί για τη λήψη αποφάσεων σε απρόβλεπτες περιστάσεις» (Stalidis et al., 2015). Η έννοια της σοφίας αποτελεί ένα

περίπλοκο και πολύ εξειδικευμένο πεδίο, το οποίο στις περισσότερες περιπτώσεις επιχειρηματικών δράσεων δεν λαμβάνεται υπόψιν.

Η εξελικτική μετάβαση των επιπέδων που περιγράφεται παραπάνω, αποτυπώνεται και στο Σχήμα[3.1], που ακολουθεί:



Σχήμα 3.1: .
Πίνακας Μεταβάσεων Μεταξύ Δεδομένων - Πληροφορίας - Γνώσης - Σοφίας
(Bellinger et al., 2004.)

Ο οριζόντιος άξονας του διαγράμματος ορίζει την ποσότητα κατανόησης, ενώ ο κάθετος άξονας ορίζει την συνδεσιμότητα των στοιχείων. Η μετάβαση από το **πρώτο** στο **δεύτερο** επίπεδο, απαιτεί την *κατανόηση* των σχέσεων που συνδέουν τα δεδομένα και παράγουν τελικά χρήσιμες πληροφορίες. Η μετάβαση από το **δεύτερο** στο **τρίτο** επίπεδο, απαιτεί την *κατανόηση* των μοτίβων που προκύπτουν μελετώντας τις πληροφορίες. Τέλος, η μετάβαση από το **τρίτο** στο **τέταρτο** επίπεδο, απαιτεί την *κατανόηση* των αρχών, για το σύνολο της γνώσης που έχει παραχθεί. Σε κάθε μεταβατική ενέργεια μεταξύ των επιπέδων, απαραίτητη προϋπόθεση αποτελεί η ύπαρξη του στοιχείου της *κατανόησης*, η οποία όμως δεν μπορεί να συμπεριλαμβάνεται ως αυτόνομο και ξεχωριστό επίπεδο ανάμεσα στα υπόλοιπα (Bellinger et al., 2004).

3.2 Επιχειρηματική Ευφυΐα

Ο σύγχρονος επιχειρηματικός κόσμος προσφέρει μεγάλες ευκαιρίες αλλά και σοβαρές προκλήσεις. Ο ισχυρός ανταγωνισμός, οι πρόσφατη οικονομική κρίση, η ανάγκη για προσαρμογή στη ραγδαία εξέλιξη των τεχνολογιών, απαιτούν συνολικά μια εύστοχη αναβάθμιση των πρακτικών που ασκεί η διοίκηση των επιχειρήσεων και εκσυγχρονισμό των διαδικασιών, που σχετίζονται με τη διαδικασία λήψης αποφάσεων. Ακόμη, για την κατάκτηση ανταγωνιστικού πλεονεκτήματος και την ανάληψη ρίσκου, η επιχείρηση καλείται να γνωρίζει σε βάθος το περιβάλλον στο οποίο ασκεί δράση και συγχρόνως τροφοδοτείται με επαρκή και έγκαιρη πληροφόρηση.

3.2.1 Ορισμός και Τυπική Εφαρμογή της Επιχειρηματικής Ευφυΐας

Τις τελευταίες δύο δεκαετίες, οι ανάγκες που αναφέρθηκαν παραπάνω θέτουν στο κέντρο του ενδιαφέροντος την Επιχειρηματική Ευφυΐα. Με τον όρο Επιχειρηματική Ευφυΐα ορίζουμε «ένα σύνολο από μεθόδους ανάλυσης, τεχνολογίες, ικανότητες και στρατηγικές, οι οποίες στόχο έχουν την επεξεργασία των διαθέσιμων δεδομένων και την εξαγωγή πληροφορίας από αυτά, για την υποστήριξη της διαδικασίας λήψης επιχειρηματικών αποφάσεων.» (Kirkos, 2015). Κατά κύριο λόγο, το πεδίο εφαρμογής της Επιχειρηματικής Ευφυΐας στηρίζεται σε μεγάλες αποθήκες ποιοτικών δεδομένων και αποσκοπεί στην διάδοση της πληροφορίας ή της γνώσης σε ολόκληρο τον οργανισμό, τόσο σε επίπεδα τακτικής και στρατηγικής, όσο και στο επίπεδο της επιχειρηματικής της δράσης.

Μια τυπική εφαρμογή Επιχειρηματικής Ευφυΐας αποτελείται από πέντε βασικά στάδια (Li Niu et al., 2009, σ. 19-29):

1. **Παροχή Δεδομένων:** Οι πηγές από τις οποίες μπορούν τα συστήματα Επιχειρηματικής Ευφυΐας να αντλήσουν δεδομένα είναι πολυποίκιλες. Μάλιστα, κάθε μία από τις πηγές αυτές μπορεί να αντιπροσωπεύει και ένα διαφορετικό τμήμα ή μονάδα της επιχείρησης (τμήμα ανθρωπίνων πόρων, οικονομικών, παραγωγής, μάρκετινγκ). Με το πέρας αυτής της διαδικασίας, τα δεδομένα που συλλέγονται οδηγούνται προς επεξεργασία, οργάνωση και μετασχηματισμό με στόχο την ευκολότερη διαχείριση και ανάλυσή τους.
2. **Ανάλυση Δεδομένων:** Στο στάδιο αυτό κάνοντας χρήση τεχνικών μοντελοποίησης, εξόρυξης δεδομένων και οπτικοποίησης, μετατρέπουμε το σύνολο των δεδομένων σε πληροφορίες ή γνώση. Στο στάδιο, αυτό παρέχεται στη διοίκηση καλύτερη εικόνα του επιχειρηματικού της περιβάλλοντος και δίνεται η δυνατότητα να βελτιώνει την ποιότητα και την αποτελεσματικότητα των αποφάσεων που λαμβάνονται.
3. **Επίγνωση της κατάστασης:** Βασικό στάδιο κατά το οποίο η διοίκηση των εταιρειών αποκτά βαθύτερη κατανόηση της τρέχουσας κατάστασης, γύρω από τις κρίσιμες αποφάσεις τις οποίες καλείται να πάρει. Τα συστήματα Επιχειρηματικής Ευφυΐας πρέπει να ενισχύουν αισθητά τον βαθμό επίγνωσης και αντίληψης της κατάστασης και να παρέχουν σχέδια και εναλλακτικές προς ενίσχυση της διαδικασίας λήψης αποφάσεων.
4. **Εκτίμηση Κινδύνου:** Ενισχύοντας το στάδιο εκτίμησης κινδύνου μια επιχείρηση πραγματοποιεί με μεγαλύτερη σιγουριά εκτιμήσεις για το μέλλον, αντιλαμβάνεται έγκαιρα κινδύνους και ανταγωνιστικές απειλές, δρώντας προνοητικά και αποφασιστικά.
5. **Υποστήριξη Αποφάσεων:** Το τελικό στάδιο αφορά αποκλειστικά την διαδικασία λήψης κατάλληλων αποφάσεων. Εξάλλου, η αφορμή που οδήγησε στη δημιουργία των συστημάτων επιχειρηματικής ευφυΐας δεν ήταν παρά η ανάγκη λήψης

αποφάσεων με σύνεση, υπο την ενδελεχή εξέταση των διαθέσιμων δεδομένων που μια επιχείρηση έχει υπό την κατοχή της.

3.2.2 Η αρχιτεκτονική ενός συστήματος Επιχειρηματικής Ευφυΐας

Η λειτουργία των συστημάτων της Επιχειρηματικής Ευφυΐας βασίζεται σε μια αλληλουχία διεργασιών και φάσεων με βασικό στόχο την βέλτιστη λήψη αποφάσεων. Για την αναπαράσταση των συστημάτων της Επιχειρηματικής Ευφυΐας γίνεται χρήση μιας πυραμίδας [Σχήμα 3.2], με διαβαθμισμένα επίπεδα, όπου σε καθένα εκ των οποίων αντιστοιχεί και ένα σύστημά της. Με τη σταδιακή ανέλιξη μεταξύ των βαθμίδων, είναι εμφανές ότι αυξάνεται και η ποιότητα της λήψης των επιχειρηματικών αποφάσεων (Kirkos, 2015).

- **Επίπεδο 0 - Πηγές Δεδομένων :** Τα δεδομένα αποτελούν, όπως προαναφέρθηκε, το βασικό υλικό μιας διαδικασίας λήψης αποφάσεων. Το κατώτατο επίπεδο της πυραμίδας συνεπώς περιλαμβάνει, σε ανεπεξέργαστη μορφή, όλα εκείνα τα δεδομένα που συλλέγει σχολαστικά μία επιχείρηση. Βασική πηγή προέλευσης των δεδομένων αποτελούν τα συστήματα ERP, SCM, CRM, οι βάσεις δεδομένων και τα έγγραφα εντός των εταιρειών, όπως επίσης δεδομένα τρίτων παρόχων, εκτός των οργανισμών (Kirkos, 2021). Μεμονωμένα, τα δεδομένα δεν εξυπηρετούν τη διαδικασία λήψης επιχειρηματικών αποφάσεων. Η βασική αιτία εμπίπτει στο γεγονός ότι, τα δεδομένα παρέχουν πληροφορίες γύρω από συγκεκριμένα περιστατικά, δεν αποτυπώνουν συμπεράσματα μια ευρύτερης οικονομικής ύφεσης, ανάπτυξης ή καταναλωτικής συμπεριφοράς, είναι αταξινόμητα και η μορφή στην οποία τελικά αποθηκεύονται, ποικίλλει. Συνεπώς, η επεξεργασία και οργάνωση των δεδομένων αποτελεί και το ενακτήριο λάκτισμα για την εφαρμογή συστημάτων επιχειρηματικής ευφυΐας που με τη σειρά τους θα οδηγήσουν σε βελτιωμένη λήψη αποφάσεων.

- **Επίπεδο 1 - Αποθήκες Δεδομένων :** Σε αυτό το επίπεδο αποθηκεύονται τα δεδομένα υλοποιώντας κυρίως παραδοσιακά σχεσιακά συστήματα διαχείρισης βάσεων δεδομένων. Οι εργασίες που εκτελούνται επί των δεδομένων είναι η εξαγωγή τους, ο μετασχηματισμός τους, σύμφωνα με ένα σύνολο κανόνων που έχουν ορισθεί και τελικά η φόρτωση των δεδομένων, σε μία κεντρική αποθήκη δεδομένων (Extract, Transform, Load - ETL).

Ο βέλτιστος σχεδιασμός και έπειτα η ανάπτυξη ενός αποτελεσματικού ETL αποτελεί προϋπόθεση υψηλής σημασίας. Βασική απόρροια ενός αποδοτικού ETL είναι η ευελξία ολόκληρου του συστήματος, η αυξημένη ταχύτητα επεξεργασίας και η υλοποίηση βάσεων δεδομένων υψηλής ποιότητας. Τα σημερινά συστήματα αποθηκών δεδομένων, μπορούν να έχουν ευρύ εταιρικό πεδίο και τα προς αποθήκευση δεδομένα μπορούν να ενημερώνονται σε πραγματικό χρόνο.

- **Επίπεδο 2 - Διερεύνηση Δεδομένων :** Το στάδιο της διερεύνησης επικεντρώνεται στην λήψη απαντήσεων - αποτελεσμάτων, θέτοντας πρωτίστως ερωτήματα, γνωστά και ως «queries» γύρω από μία βάση δεδομένων. Ανάλογα με τις τελικές απαντήσεις που λαμβάνει ο χρήστης συντάσσει και τις πρώτες αναφορές, οι οποίες μάλιστα μπορούν να έχουν και οπτικό χαρακτήρα, υπό τη μορφή γραφήματος ή πίνακα. Με αυτό το τρόπο επιταχύνεται ο χρόνος κατανόησης μίας παρουσίασης αποτελεσμάτων και γίνεται ευκολότερη και πιο ευνόητη η διαδικασία συνολικά. Το επίπεδο της διερεύνησης επιδέχεται και στατιστική ανάλυση των βάσεων δεδομένων και ο βασικός σκοπός σε αυτή τη βαθμίδα της πυραμίδας είναι να τεθούν υποθέσεις από τον χρήστη-αναλυτή και τα αποτελέσματα με τη σειρά του να επιβεβαιώσουν ή να διαψεύσουν τις υποθέσεις αυτές (Kirkos, 2021).

- **Επίπεδο 3 - Εξόρυξη Δεδομένων :** Η Εξόρυξη Δεδομένων είναι ένα διεπιστημονικό πεδίο και έχει απασχολήσει τόσο τη βιομηχανία όσο και τον ακαδημαϊκό κόσμο. Κυρίαρχος σκοπός της έρευνας και ανάπτυξης των μεθόδων εξόρυξης

δεδομένων είναι ο εντοπισμός επαναλαμβανόμενων μοτίβων ή τάσεων, που προκύπτουν μέσα από ένα πλήθος δεδομένων μίας βάσης δεδομένων. Η εξόρυξη δεδομένων δεν συλλέγει και διαχειρίζεται απλώς μεγάλες ποσότητες δεδομένων, επιδιώκει μια βαθύτερη ανάλυση και πρόβλεψη. Τα κύρια εργαλεία της διαδικασίας αυτής προέρχονται από τους επιστημονικούς κλάδους της Στατιστικής, της Τεχνητής Νοημοσύνης και της Μηχανικής Μάθησης. Μία αξιοσημείωτη διαφορά του επιπέδου - Διερεύνησης Δεδομένων με αυτό της Εξόρυξης Δεδομένων είναι η έλλειψη αναγκαιότητας κατάθεσης αρχικών υποθέσεων στο σύστημα. Οι αλγόριθμοι που εφαρμόζονται, παράγουν άμεσα πληροφορίες ή γνώση, μέσω της επεξεργασίας των δεδομένων που τροφοδοτείται στο σύστημα. Το τελικό προϊόν, συνήθως, είναι ένα μοντέλο που υπόκειται σε συγκεκριμένους κανόνες και στη συνέχεια διακρίνεται ανάμεσα σε δυο τύπους μοντέλων (Verykios et al., 2015):

- Ο πρώτος τύπος μοντέλου ονομάζεται *μοντέλο πρόβλεψης* (predictive) και προβλέπει τιμές για κάποιο χαρακτηριστικό υψηλού ενδιαφέροντος, το οποίο συχνά μάλιστα εξαρτάται από τη συμπεριφορά άλλων χαρακτηριστικών.
- Ο δεύτερος τύπος μοντέλου ονομάζεται *περιγραφικό μοντέλο* (descriptive). Σε αυτή την περίπτωση, βασικός σκοπός είναι ο εντοπισμός προτύπων ή σχέσεων ανάμεσα στα δεδομένα, ώστε να αναδειχθούν τα αίτια που προκαλούν συγκεκριμένα μοτίβα συμπεριφοράς των δεδομένων μίας βάσης.

- **Επίπεδο 4 - Βελτιστοποίηση** : Τα συστήματα της Επιχειρηματικής Ευφυΐας των κατώτερων βαθμίδων της πυραμίδας, με την άφιξη στο επίπεδο της βελτιστοποίησης, έχουν ήδη παράξει ένα σύνολο πληροφοριών και γνώσης. Στη συνέχεια, παρουσιάζεται ένα σύνολο πιθανών λύσεων στήριξης επιχειρηματικών αποφάσεων, από τις οποίες η διοίκηση καλείται να επιλέξει. Εξετάζοντας τις διεργασίες του επιπέδου, ως προς τα είδη των λύσεων, προκύπτουν τρεις κατηγορίες προβλημάτων (Kirkos, 2021):

- Η πρώτη κατηγορία αποτελείται από τα διχοτόμα προβλήματα, στα οποία αναλογούν δύο μόνο πιθανές λύσεις.
- Η δεύτερη κατηγορία αποτελείται από τα προβλήματα πολλαπλών λύσεων και ο αριθμός των προκειμένων λύσεων είναι πεπερασμένος.
- Η τρίτη κατηγορία αφορά προβλήματα απεριόριστου αριθμού λύσεων. Στην περίπτωση αυτή, αναζητείται η πλέον βέλτιστη λύση.

- **Επίπεδο 5 - Λήψη Αποφάσεων :** Το στάδιο αυτό είναι άμεσα συνδεδεμένο με τον ανθρώπινο παράγοντα, καθώς στο τέλος είναι αυτός που καλείται να οριστικοποιήσει την απόφαση που πρόκειται να λάβει. Συγκεκριμένα, οι ανθρώπινες εμπειρίες, γνώσεις και η λογική λειτουργούν συνεργατικά με τις πληροφορίες που παράχθηκαν, από τις διεργασίες των συστημάτων της Επιχειρηματικής Ευφυΐας και λειτουργούν εξίσου ενισχυτικά, στην ποιότητα της τελικής απόφασης (ο.π.).



Σχήμα 3.2: .
Πυραμίδα Συστημάτων Επιχειρηματικής Ευφυΐας (Kirkos, 2015).

3.3 Βασικές Τεχνικές Εξόρυξης Πληροφοριών και Γνώσης

Στο περιβάλλον των επιχειρήσεων η Εξόρυξη Δεδομένων χρησιμοποιείται κυρίως, για (Ahmed, 2004, σ. 255-259):

- διατήρηση και απόκτηση πελατών
- διερεύνηση, εντοπισμό και διόρθωση αναποτελεσματικών μεθόδων
- μείωση της απάτης
- χαρτογράφηση ανεξερευνήτων πτυχών του διαδικτύου

Το πλήθος των μεθόδων εξόρυξης δεδομένων είναι μεγάλο. Με βάση το είδος των δεδομένων και της γνώσης που προκύπτει, ως προϊόν της διαδικασίας εξόρυξης, οι μέθοδοι κατηγοριοποιούνται και επιλέγονται ανάλογα με τις ανάγκες του συστήματος.

3.3.1 Κατηγοριοποίηση

Όπως υποδεικνύει η μέθοδος, τα δεδομένα πρέπει να υπάγονται σε κατηγορίες. Η μέθοδος της Κατηγοριοποίησης (classification), θεωρείται από τις πιο διαδεδομένες μεθόδους και πρόκειται για εργασία επιβλεπόμενης μάθησης. Συγκεκριμένα, «είναι η μάθηση μιας συνάρτησης, η οποία απεικονίζει ένα αντικείμενο (συνήθως αναπαρίσταται ως ένα διάνυσμα τιμών για τις χαρακτηριστικές του ιδιότητες) σε μία τιμή μιας κατηγορικής μεταβλητής, η οποία είναι γνωστή και ως κλάση (ή κατηγορία)».

Το τελικό αποτέλεσμα μιας διαδικασίας κατηγοριοποίησης είναι μια κλάση με δείγματα και οι τιμές τις οποίες λαμβάνει είναι από ένα πεπρασμένο διακριτό σύνολο. Η τεχνική της μεθόδου είναι ικανή να επεξεργαστεί μια ευρύτερη ποικιλία δεδομένων από τη μέθοδο της παλινδρόμησης και αυξάνεται γι αυτόν τον λόγο η δημοτικότητά της (Verykios et al., 2015).

3.3.2 Παλινδρόμηση

Η παλινδρόμηση είναι μία μέθοδος που φέρει κοινά στοιχεία με την μέθοδο της κατηγοριοποίησης. Πρόκειται για εργασία επιβλεπόμενης μάθησης και στόχος της είναι η «μάθηση» ή αλλιώς η «εκπαίδευση» (training) μιας συνάρτησης, η οποία απεικονίζει ένα αντικείμενο σε μία πραγματική μεταβλητή» (ο.π.). Μία ειδοποιός διαφορά μεταξύ των μεθόδων Κατηγοριοποίησης - Παλινδρόμησης είναι ότι στην περίπτωση της δεύτερης υπολογίζονται αριθμητικές τιμές. Χρήση της μεθόδου στον χώρο των επιχειρήσεων συναντάμε σε περιπτώσεις πρόβλεψης αριθμητικών τιμών (κέρδη , ύψη πωλήσεων, κλπ) (Kirkos, 2015).

3.3.3 Ανάλυση Χρονοσειρών

Η εξέλιξη κάποιων μεγεθών, σε ορισμένες περιπτώσεις, συνάδει με την πάροδο του χρόνου. Με την χρήση χρονοσειρών, σημείων δηλαδή μιας ακολουθίας που καταγράφονται από μετρήσεις που πραγματοποιούνται, επιτυγχάνεται η αναπαράσταση της εξέλιξης του μεγέθους, στη διάρκεια του χρόνου. Με την μέθοδο Ανάλυσης Χρονοσειρών, δίνεται η δυνατότητα συλλογής συμπερασμάτων από διαφορετικά χρονικά διαστήματα και αυτή η προσέγγιση αποδίδει υψηλής σημασίας συμπεράσματα σχετικά με το φαινόμενο που εξετάζεται (ο.π.).

3.3.4 Συσταδοποίηση

Στη μέθοδο της συσταδοποίησης κύριος σκοπός είναι, «η δημιουργία συστάδων (clusters), δηλαδή ομάδων, οι οποίες θα περιέχουν όμοια ή παρεμφερή δείγματα. «Ουσιαστικά αναζητείται ένα πεπερασμένο σύνολο κατηγοριών ή συστάδων, για να περιγράψει τα δεδομένα» (Verykios et al., 2015). Αποτελεί μέθοδο μη επιβλεπόμενης μάθησης και διαφέρει από τις παραπάνω διεργασίες Παλινδρόμησης και Κατηγοριοποίησης. Η βασική διαφορά εντοπίζεται στο γεγονός ότι, οι κατηγορίες που σχηματίζει η μέθοδος της συσταδοποι-

ήσης δεν είναι καταχωρημένες και γνωστές από την έναρξη της διεργασίας. Αντιθέτως, ο αλγοριθμός αναζητά και σχηματίζει ομάδες στοιχείων που μοιάζουν μεταξύ τους και φέρουν κοινά χαρακτηριστικά. Ο σχηματισμός των συστάδων πρέπει να γίνεται με αυστηρό τρόπο, τέτοιο ώστε τα στοιχεία μίας συστάδας να μην ανήκουν συγχρόνως και σε άλλη. Με το πέρας του σχηματισμού συστάδων, ακολουθεί η διαδικασία δημιουργίας κανόνων που περιγράφουν κάθε ομάδα (Kirkos, 2015).

Ονομαστικά βασικές μέθοδοι συσταδοποίησης, είναι:

- Κανόνες Συσχέτισης
- Μέθοδοι Βασισμένες σε Μοντέλα
- Ιεραρχικές Μέθοδοι
- Διαχωριστικές Μέθοδοι

3.3.5 Αλγόριθμος K-means

Το έτος 1967 ο James MacQueen (1967, σ. 281-297) εισήγαγε τον όρο K-Means στο χώρο των επιστημών. Είναι μια διαιρετική μέθοδος ανάλυσης συστάδων και ανήκει στην κατηγορία αλγορίθμων μάθησης χωρίς επίβλεψη. Στόχος της μεθόδου είναι να ομαδοποιήσει το σύνολο των δεδομένων σε έναν αριθμό συστάδων, που έχει προκαθορίσει ο χρήστης, με τρόπο τέτοιο ώστε τα επιμέρους στοιχεία κάθε υποσυνόλου να φέρουν μεγάλη ομοιότητα μεταξύ τους. Ο αλγόριθμος K-Means μπορεί να υλοποιηθεί τόσο ποσοτικά όσο και ποιοτικά. Στην πρώτη περίπτωση, γίνεται χρήση αριθμητικών μεταβλητών, ενώ στη δεύτερη περίπτωση χρησιμοποιούνται κατηγορικές μεταβλητές.

Στη συνέχεια, δίνεται η αλγοριθμική περιγραφή της μεθόδου βήμα προς βήμα (Verykios et al., 2015):

1. Αυθαίρετη επιλογή k τυχαίων σημείων, τα οποία ονομάζονται κεντροειδή centroids των συστάδων. Ο αριθμός k , υποδηλώνει το συνολό των συστάδων που θα δημιουργηθούν κατά την εκτέλεση του αλγορίθμου.
2. Στη συνέχεια ο αλγόριθμος εκτελεί επαναληπτικά δύο βήματα.
 - Το πρώτο βήμα αφορά την ανάθεση, ενός στοιχείου από το σύνολο των δεδομένων, σε κάποια συστάδα
 - Το δεύτερο βήμα αφορά τον επαναπροσδιορισμό και τη μετατόπιση του κεντροειδούς κάθε συστάδας.

Πιο συγκεκριμένα, κατά την εκτέλεση του πρώτου βήματος, ο αλγόριθμος K-Means εξετάζει την απόσταση κάθε στοιχείου του συνόλου, με τα κεντροειδή των συστάδων. Τελικά, το δείγμα κατατάσσεται σε συστάδα της οποίας η απόσταση κεντροειδούς-στοιχείου είναι η μικρότερη. Ο υπολογισμός των αποστάσεων παραπάνω, συνήθως, υπολογίζεται κάνοντας χρήση της Ευκλείδειας απόστασης (Kirkos, 2015).

Στο δεύτερο βήμα, υπολογίζεται ο μέσος όρος των δειγμάτων κάθε συστάδας. Έπτερα, γίνεται επαναπροσδιορισμός των κέντρων βάρους κάθε υποσυνόλου, με σκοπό να είναι πιο αντιπροσωπευτικό στην πρόσφατα διαμορφωμένη συστάδα. Η διαδικασία παύει να επαναλαμβάνεται, όταν πλέον οι μετατοπίσεις των στοιχείων είναι ασήμαντες και δεν ξεπερνούν μία ελάχιστη τιμή αποστάσεως, την οποία ορίζει ο χρήστης ως κατώφλι.

Η πολυπλοκότητα του K-Means είναι μικρή και αυτό τον καθιστά από τους πλέον διαδεδομένους αλγορίθμους στο περιβάλλον της Εξόρυξης Δεδομένων και της Επιχειρηματικής Ευφυΐας γενικότερα. Η απλότητα των βημάτων διεκπεραίωσης διευκολύνει και επιταχύνει τη διαδικασία συσταδοποίησης, ακόμη και σε πολύ μεγάλα σύνολα δεδομένων.

Στον αντίποδα, ο αριθμός των συστάδων k πρέπει να προκαθορίζεται από τον χρήστη και η ποιότητα του τελικού αποτελέσματος εξαρτάται άμεσα από την επιλογή αυτή. Κάθε διαφορετική εκχώρηση συστάδων μπορεί να οδηγήσει σε αισθητά διαφορετικό αποτέλεσμα. Τέλος, ο αλγόριθμος K-Means παρουσιάζει μεγάλη ευαισθησία σε ακραίες τιμές,

επηρεάζοντας άμεσα τους επανυπολογισμούς των κέντρων και κατ' επέκταση την συνολική διαδικασία (ο.π.π.)

3.4 Μειονεκτήματα - Πλεονεκτήματα Επιχειρηματικής Ευφυΐας

Η εφαρμογή συστημάτων Επιχειρηματικής Ευφυΐας αποτελεί ένα ισχυρό εργαλείο στο χώρο των επιχειρήσεων. Η απόφαση της διοίκησης των εταιρειών, να χρησιμοποιήσει νέες μεθόδους και διαδικασίες που υπάγονται στο περιβάλλον της Επιχειρηματικής Ευφυΐας, εγχυμονεί τις περισσότερες φορές, ορισμένους κινδύνους. Στη συνέχεια, γίνεται μια εκτενής προσπάθεια ανάδειξης των μειονεκτημάτων και πλεονεκτημάτων που παρατηρούνται, κατα την αξιοποίηση των συστημάτων αυτών.

3.4.1 Μειονεκτήματα και Περιορισμοί

- Υψηλό Κόστος Υλικού: Ένας από τους συνηθέστερους περιορισμούς χρήσης συστημάτων Επιχειρηματικής Ευφυΐας είναι η έλλειψη δεδομένων. Η συλλογή και συντήρηση δεδομένων ή σε άλλες περιπτώσεις, η αγορά δεδομένων απο τρίτους φορείς, μπορεί να αυξήσει αισθητά το κόστος για έναν οργανισμό (Negash, 2004).
- Ανθρώπινοι Πόροι - Εκπαιδευτικά Κενά: Η αδυναμία ανταπόκρισης του ανθρώπινου δυναμικού των εταιρειών στα νέα πληροφοριακά συστήματα της Επιχειρηματικής Ευφυΐας, δημιουργεί επιτακτικά την ανάγκη για αναβάθμιση των γνώσεων και δεξιοτήτων (upskilling - reskilling), γεγονός που επιφέρει μεγάλη οικονομική επιβάρυνση καθώς και χρονική επιβράδυνση στην εξέλιξη των έργων μέχρι την ολοκλήρωση της κατάρτισης των εργαζομένων. Εναλλακτικά, προσλήψεις νέων εργαζομένων στον οργανισμό αποτελούν επιπλέον οικονομική επιβάρυνση και η ένταξή τους επιφέρει χρονική καθυστέρηση, αντίστοιχα.

- Ασυμβατότητα Πληροφοριακών Συστημάτων: Οι εταιρείες επιστρατεύουν διάφορους τρόπους ώστε να ενισχύσουν τη δράση τους και να αντιμετωπίσουν τον ανταγωνισμό. Η χρήση πολύπλοκων συστημάτων πληροφορικής, εντός των οργανισμών, δυσχεραίνει την ένταξη και λειτουργικότητα των συστημάτων της Επιχειρηματικής Ευφυΐας.
- Αδυναμία οριζόντιας επικοινωνίας: Μεγάλη δυσκολία παρουσιάζεται, αρκετές φορές, στην επικοινωνία των διοικητικών στελεχών μιας επιχείρησης, μεταξύ τους. Το πρόβλημα μάλιστα, φαίνεται να εντείνεται περαιτέρω στη συνεργασία των υπόλοιπων τμημάτων με το εξειδικευμένο τμήμα εργαζομένων στην Επιχειρηματική Ευφυΐα, τους οποίους χρειάζεται να κατευθύνουν τακτικά με βάση τις αποφάσεις που λαμβάνουν (Stalidis et al., 2015).

3.4.2 Πλεονεκτήματα και Οφέλη

- Αξιοπιστία και Εγκυρότητα: Οι τεχνολογίες που αξιοποιούνται προσφέρουν ουσιαστική πληροφόρηση. Τα βήματα των μεθόδων που ακολουθούν τα συστήματα της Επιχειρηματικής Ευφυΐας είναι αξιόπιστα και έγκυρα.
- Εντοπίζει τους παράγοντες που οδηγούν στην κερδοφορία: Ανακαλύπτει και αξιολογεί τη σχέση προϊόν-πελάτη και προτείνει στην επιχείρηση τρόπους, που στοχεύουν στην βελτίωση της σχέσης της με τον καταναλωτή. Η επιχείρηση ακολουθώντας την κατεύθυνση αυτή, επιτυγχάνει αύξηση της εμπιστοσύνης και της προτίμησης του καταναλωτικού της κοινού και της συνολικής της κερδοφορίας ταυτόχρονα (Jaklič, 2008).
- Πρόβλεψη - Ανάλυση - Δράση: Τα μοντέλα που αξιοποιούν τα συστήματα της Επιχειρηματικής Ευφυΐας επιτρέπουν την ανασκόπηση παρελθοντικών γεγονότων. Μέσω προσεκτικής ανάλυσης και χρήσης μοντέλων πρόβλεψης, γίνεται εκτίμηση μελλοντικών τάσεων της αγοράς και των ευρύτερων καταναλωτικών συμπεριφο-

ρών. Το γεγονός αυτό, παρέχει ισχυρό ανταγωνιστικό προβάδισμα στην επιχείρηση και ώριμη διαχείριση κινδύνου (Siegel, 2013).

- Η διαδικασία λήψης επιχειρηματικών αποφάσεων που πραγματοποιεί η διοίκηση ενός οργανισμού, χρησιμοποιώντας ποιοτική πληροφόρηση και γνώση που της παρέχουν τα συστήματα της Επιχειρηματικής Ευφυΐας, προσεγγίζει τη βέλτιστη εκδοχή της.

Κεφάλαιο 4

Οπτικοποίηση Δεδομένων

Η παραγωγή δεδομένων, στους σύγχρονους οργανισμούς, είναι μια αδιάκοπη διαδικασία που προκύπτει καθολικά, από όλο το εύρος των επιχειρήσεων, από τα διοικητικά στελέχη, στις ομάδες αναλυτών, στα τηλεφωνικά κέντρα έως και στους υπαλλήλους χειρονακτικών εργασιών στις γραμμές παραγωγής εργοστασίων. Ο τεράστιος όγκος Μεγάλων Δεδομένων που συσσωρεύεται, απαιτεί τις κατάλληλες τεχνολογίες και μεθόδους επεξεργασίας, ώστε να μπορεί να αξιοποιηθεί κατάλληλα και ύστερα τα συμπεράσματα που προκύπτουν, να γίνονται άμεσα κατανοητά και ωφέλιμα. Με τον όρο της Οπτικοποίησης Δεδομένων, αναφερόμαστε στην παρουσίαση δεδομένων σε γραφική ή εικονογραφική διάταξη (Ajibade et al., 2016, σ. 105-133). Τα εργαλεία της Οπτικοποίησης Δεδομένων επιτρέπουν δυσνόητα αποτελέσματα αναλύσεων, περίπλοκες ιδέες και συμπεράσματα, να μετατρέπονται σε ευνόητες οπτικές συσχετίσεις και μοτίβα. Οι σύγχρονες τεχνολογικές πλατφόρμες και λογισμικά οπτικοποίησης διευκολύνουν, σε μεγάλο βαθμό, την επεξεργασία και τις δοκιμές που επιθυμεί ο χρήστης να εφαρμόσει στο σύνολο των δεδομένων που τον αφορά. Ο χρόνος διεκπεραίωσης υπολογισμών και οπτικής αναπαράστασης των αποτελεσμάτων, έχει μειωθεί δραματικά και ο προσανατολισμός των αναλύσεων προσαρμόζεται ευκολότερα πλέον στις προτιμήσεις των αναλυτών.

Μέσω της οπτικοποίησης των δεδομένων επιτυγχάνεται η απεικόνιση σχέσεων συνάφειας, η διασπορά συμβάντων, η άνοδος και κάθοδος τάσεων, γίνονται συγκρίσεις τιμών και παρουσιάζονται πολλές ακόμη χρήσιμες πληροφορίες. Τέλος, η γραφική αναπαράσταση, εμπεριέχει και το στοιχείο μιας διαδικασίας παρουσίασης πληροφοριών υψηλότερης αισθητικής αξίας, η οποία γίνεται ευκολότερα αποδεκτή από τον ανθρώπινο, σε αντίθεση με περιπτώσεις μεγάλων δυσνόητων κειμένων που απαιτούν ενέργεια και χρόνο για ανάγνωση και πόσο μάλλον κατανόηση. Η δυνατότητα της Οπτικοποίησης Δεδομένων να διασφαλίζει και να επιταχύνει τη διαδικασία εμπέδωσης περίπλοκων ιδεών, θέτει νέους ορίζοντες για την ανάλυση των Μεγάλων Δεδομένων και ενισχύει την αποτελεσματικότητα της διαδικασίας λήψης αποφάσεων (Kirkos, 2015).

4.1 Ταξινόμηση Μεθόδων Απεικόνισης Δεδομένων

Το πλήθος των μεθόδων Οπτικής Αναπαράστασης Δεδομένων είναι αξιοσημείωτα υψηλό και οι διαφορές μεταξύ των μεθόδων ποικίλουν. Συνεπώς, η επιστημονική κοινότητα στην προσπάθειά της να συστηματοποιήσει το πεδίο απεικόνισης των δεδομένων, αποφάσισε να κατηγοριοποιήσει τις μεθόδους σε ειδικές κατηγορίες.

Στην ενότητα αυτή, αναλύεται η μέθοδος ταξινόμησης που ανέπτυξαν οι Kriegel και Keim (Ankerst et al., 1996). Η κεντρική ιδέα, στην οποία στηρίζεται η κατηγοριοποίηση των μεθόδων, είναι η ταξινόμηση των τεχνικών οπτικοποίησης με βάση τα εξής κριτήρια (Kirkos, 2015):

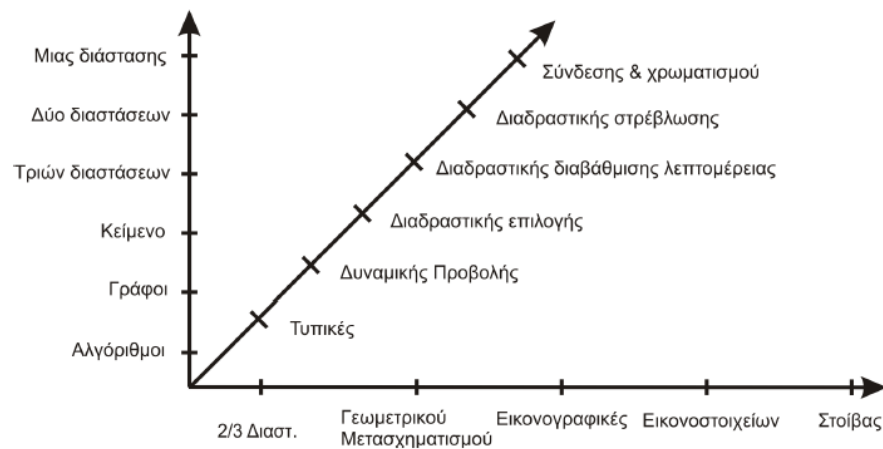
1. Τύπος Δεδομένων
2. Τεχνική Οπτικοποίησης
3. Τεχνική Αλληλεπίδρασης & Στρέβλωση

Στη συνέχεια αναφέρονται ονομαστικά τα κριτήρια που εμπεριέχονται σε κάθε κριτήριο ταξινόμησης των Kriegel και Keim.

- Ως προς τον τύπο τους τα δεδομένα μπορεί να είναι:
 1. Μονοδιάστατα
 2. Δυσδιάστατα
 3. Πολυδιάστατα
 4. Κείμενο
 5. Ιεραρχίες και Γράφοι
 6. Αλγόριθμοι
- Ως προς την τεχνική οπτικοποίησής τους τα δεδομένα μπορεί είναι:
 1. Δυσδιάστατα ή Τρισδιάστατα (2D - 3D)
 2. Γεωμετρικού Μετασχηματισμού
 3. Εικονογραφικές τεχνικές
 4. Τεχνικές Εικονοστοιχείων
 5. Τεχνικές ιεραρχικές ή στίβας.
- Ως προς την τεχνική αλληλεπίδρασης και στρέβλωσης τα δεδομένα μπορεί να είναι:
 1. Δυναμικής Προβολής
 2. Διαδραστικής Επιλογής
 3. Διαδραστικής Διαβάθμισης Λεπτομέρειας
 4. Διαδραστικής Στρέβλωσης
 5. Διαδραστικής Σύνδεσης και Χρωματισμού

Σημειώνεται πως για ένα οποιοδήποτε σύστημα, τα κριτήρια ταξινόμησης μπορούν να συνδυαστούν ελεύθερα, χωρίς επιπλέον περιορισμούς. Συγκεκριμένα, κάθε τύπος δεδομένων, έχει τη δυνατότητα να συνδυαστεί με κάθε τεχνική οπτικοποίησης και ύστερα να συνδυαστεί με κάθε τεχνική αλληλεπίδρασης και στρέβλωσης.

Στη συνέχεια, ακολουθεί το διάγραμμα ταξινόμησης, Σχήμα [4.1], όπου κάθε κριτήριο των Kriegel και Keim αποδίδεται σε έναν από τους τρεις άξονες.



Σχήμα 4.1: .

Διάγραμμα Ταξινόμησης Τεχνικών Οπτικοποίησης Δεδομένων. (Kirkos, 2015, σ. 107)

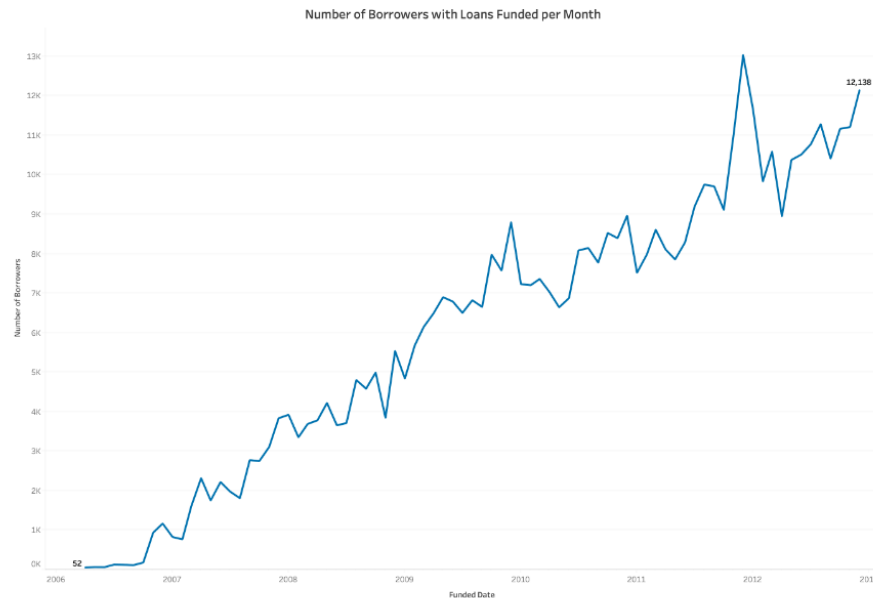
4.2 Τεχνικές Οπτικοποίησης

Με τη χρήση των τεχνικών οπτικοποίησης, δίνεται η δυνατότητα αναπαράστασης και εξέτασης πολλαπλών πληροφοριών, επιτυγχάνοντας παράλληλα χρηστικότητα, διαδραστικότητα και άμεση επέμβαση του χρήστη στα αποτελέσματα της οπτικοποίησης. Οι μέθοδοι που θα παρουσιαστούν παρακάτω, χρησιμοποιούνται τακτικά στο χώρο των επιχειρήσεων επειδή διαθέτουν μηχανισμό αξιολόγησης, δηλαδή τη δυνατότητα να παρουσιάζουν την πρόοδο που σημειώνει ένας οργανισμός.

Στη συνέχεια της ενότητας αναφέρονται οι τυπικές τεχνικές απεικόνισης δεδομένων (Ajibade, 2016, σ. 105-113· Hajirahimova, 2018, σ. 65-74· SAS, 2014· Kirkos, 2015)

1. Γραμμικό Γράφημα (Line Chart)

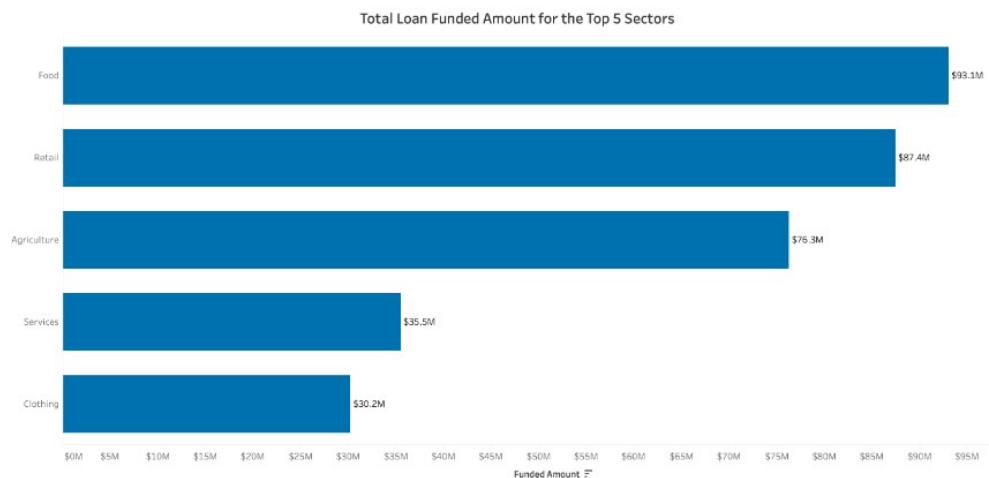
Ένα γραμμικό γράφημα, συνδέει μια σειρά σημείων δεδομένων χρησιμοποιώντας ευθύγραμμα τμήματα μεταξύ τους. Ο οριζόντιος άξονας, τις περισσότερες φορές, αντιπροσωπεύει μία διαδοχική πρόοδο τιμών και ο κάθετος άξονας περιέχει τιμές-μετρήσεις μιας μεταβλητής, για την περίοδο που εξετάζεται. Οι πιο συνηθισμένοι λόγοι για τους οποίους επιλέγονται τα γραμμικά γραφήματα, έναντι άλλων, είναι η δυνατότητά τους να παρουσιάζουν εύκολα και γρήγορα μεταβολές ή γενικότερες τάσεις μεταβλητών με την πάροδο του χρόνου ή να συγκρίνουν πλήθος μεταβλητών μεταξύ τους, για μία κοινή χρονική περίοδο.



Σχήμα 4.2: .
Παράδειγμα Γραμμικού Γραφήματος. (Tableau, 2022)

2. Ραβδόγραμμα (Bar Chart)

Τα Γραφήματα Ράβδων δίνουν τη δυνατότητα σύγκρισης αριθμητικών τιμών. Το μήκος κάθε ράβδου αντιπροσωπεύει την τιμή κάθε μεταβλητής που συγκρίνεται. Η χρήση ραβδογράμματος είναι πολύ δημοφιλής, κυρίως λόγω της απλότητας και της χρηστικότητας που το χαρακτηρίζουν.



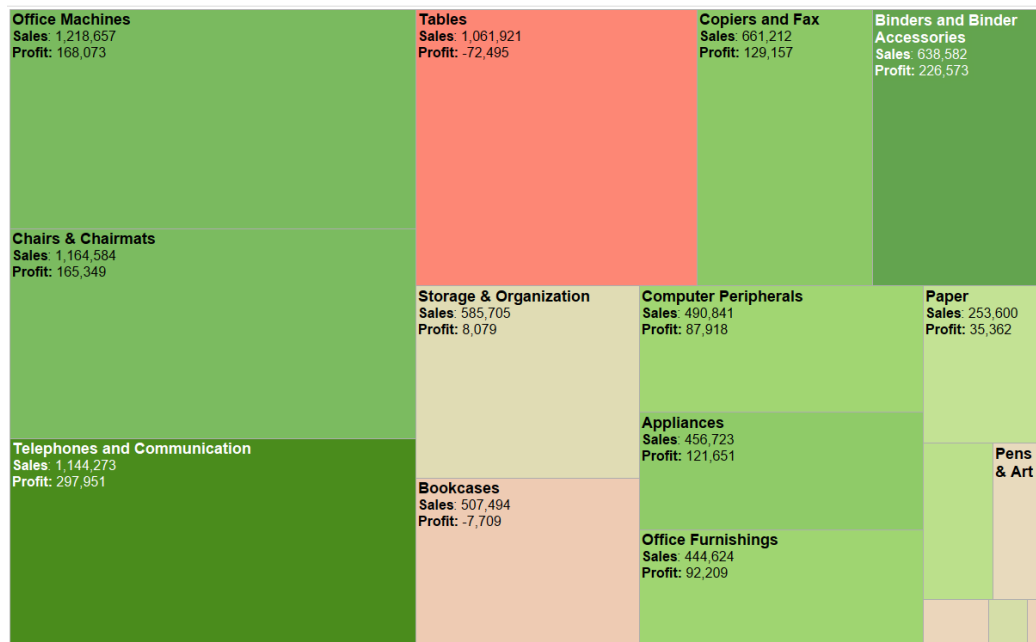
Σχήμα 4.3: .
Παράδειγμα Ραβδογράμματος. (Tableau, 2022)

3. Δενδροχάρτες (TreeMap)

Ένας δενδροχάρτης παρουσιάζει όλα τα δεδομένα ως στοιχεία ενός ιεραρχικού δένδρου. Αποτελείται από ένθετα ορθογώνια παραλληλόγραμμα που συνδέονται μεταξύ τους, όπως τα κλαδιά ενός δέντρου. Κάθε παραλληλόγραμμα διαφέρει από τα υπόλοιπα, τόσο ως προς το μέγεθός του, όσο και ως προς το χρώμα του. Η επιλογή μεγέθους και χρώματος εξαρτάται από τις δεδομένες παραμέτρους.

Η μέθοδος οπτικοποίησης με δενδροχάρτη δείχνει με ακρίβεια τις διαφορές μεταξύ των δεδομένων και το επιτυγχάνει κάνοντας χρήση μάλιστα ελάχιστου χώρου, σε σχέση με τα περισσότερα διαγράμματα δέντρων. Ένα παράδειγμα Δενδροχάρτη φαίνεται στην εικόνα που ακολουθεί [Σχήμα 4.4], όπου το μέγεθος των ορθογώνων

νίων εξαρτάται από την τιμή της μεταβλητής Sales (Πωλήσεις) και το χρώμα των αντίστοιχων ορθογωνίων εξαρτάται από τη τιμή της μεταβλητής Profit (Κέρδος).



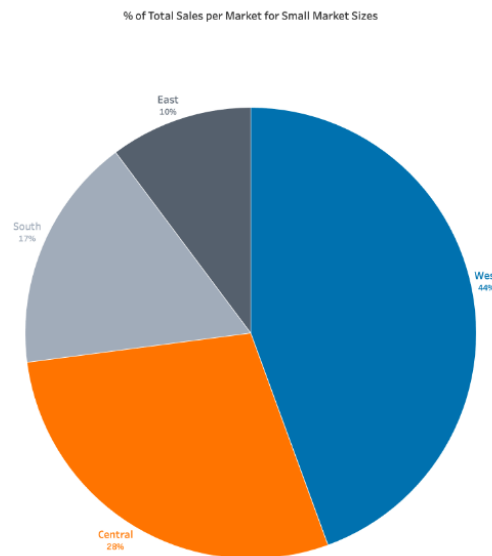
Σχήμα 4.4: .
Παράδειγμα Δενδροχάρτη. (LAB,2015)

4. Γράφημα Πίτας (Pie Chart)

Ένα Γράφημα Πίτας βοηθά στην οργάνωση και εμφάνιση δεδομένων ως ποσοστό ενός συνόλου. Ο τρόπος απεικόνισης μέσω γραφήματος πίτας είναι αρκετά δημοφιλής, όμως αρκετές φορές αμφισβητείται η εγκυρότητά του, καθώς αδυνατεί να παρουσιάσει με ευδιάκριτο τρόπο τις διαφορές μεγέθους των τμημάτων που συγκροτούν την πίτα. Γενικότερα, χρήση γραφήματος πίτας πρέπει να αποφεύγεται, όταν:

- Υπάρχουν πολλές κατηγορίες στο πρόβλημα που εξετάζεται.
- Παρόμοια ποσοστά εμφανίζονται μεταξύ διαφορετικών τιμών, εντός επιλεγμένης ιδιότητας.
- Τα ποσοστά δεν αντιπροσωπεύουν κάποιο ομοιόμορφο σύνολο και κατά συνέπεια δεν αθροίζουν στο 100 τοις εκατό.

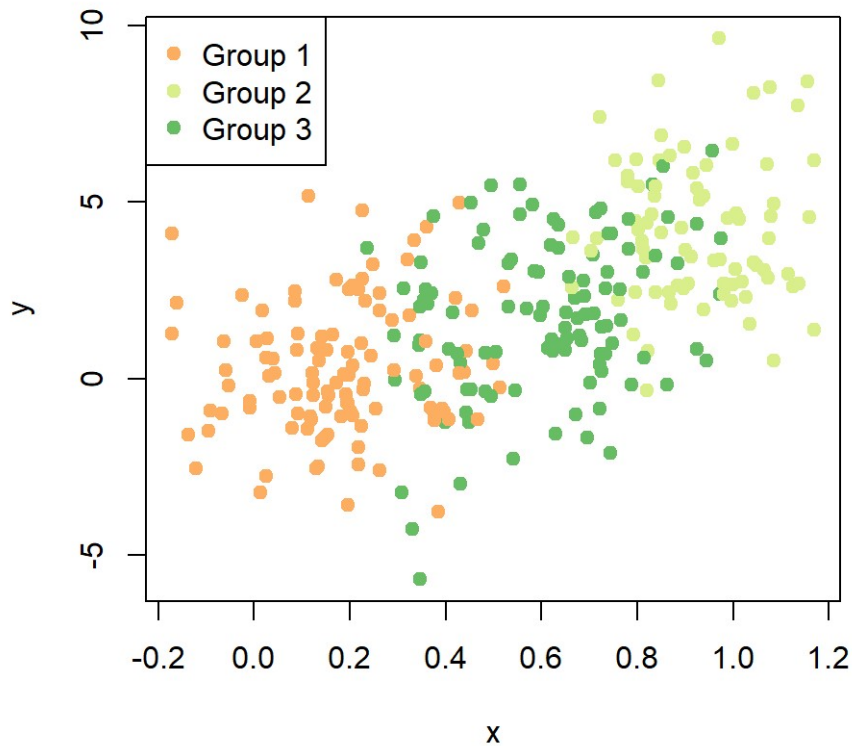
- Υπάρχουν αρνητικές τιμές.



Σχήμα 4.5: .
Παράδειγμα Γραφήματος Πίτας. (Tableau, 2022)

5. Διάγραμμα Διασποράς (Scatter Plot)

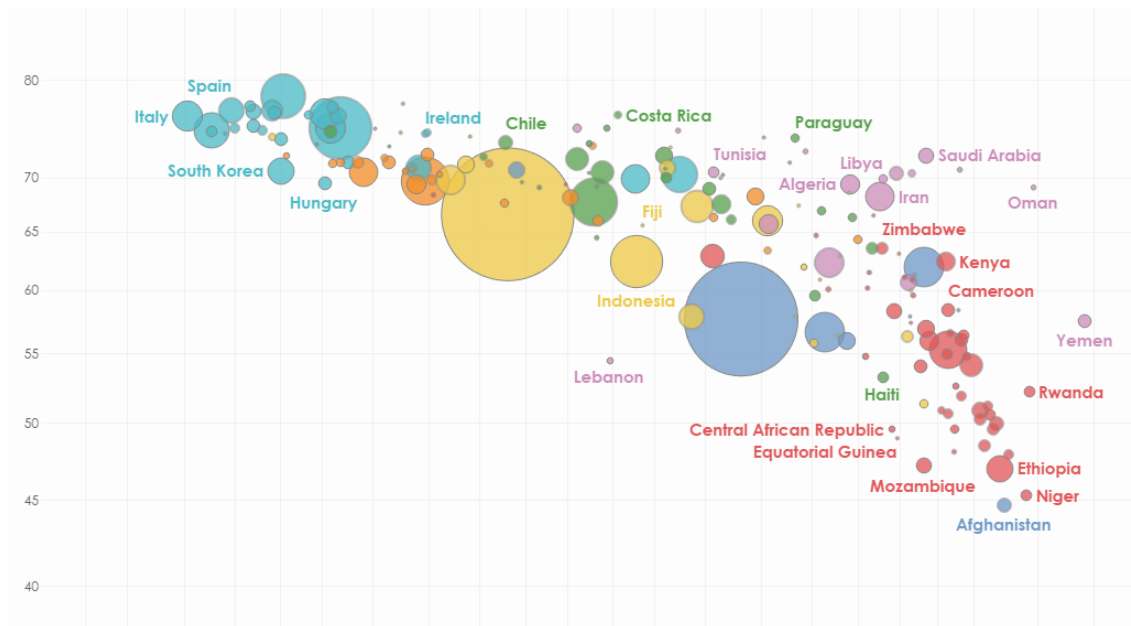
Τα Διαγράμματα Διασποράς είναι χρήσιμα για την εξέταση της σχέσης μεταξύ δύο μεταβλητών. Χρησιμοποιούν το καρτεσιανό σύστημα συντεταγμένων, πάνω στο οποίο αποτυπώνονται τα σημεία που αφορούν κάθε μεταβλητή ξεχωριστά. Δυνατότητα προσθήκης τρίτης μεταβλητής μπορεί να πραγματοποιηθεί με την προσθήκη έγχρωμων εικονιδίων στο διάγραμμα διασποράς. Το πλήθος των σημείων που παραχωρούνται στο γράφημα είναι ανάλογο του ρυθμού συσχέτισης. Δηλαδή, αύξηση του αριθμού των σημείων στο Διάγραμμα Διασποράς, σηματοδοτεί συγχρόνως και αύξηση του ρυθμού συσχέτισης των μεταβλητών που εξετάζονται.



Σχήμα 4.6: .
Παράδειγμα Διαγράμματος Διασποράς. (R-Charts,2022)

6. Διάγραμμα Φυσαλίδας (Bubble Plot)

Ένα Διάγραμμα Φυσαλίδας είναι μια παραλλαγή ενός Διαγράμματος Διασποράς, στο οποίο οι δείκτες εντός του καρτεσιανού συστήματος συντεταγμένων αντικαθίστανται από το σχήμα της φυσαλίδας. Κάθε μία από τις φυσαλίδες του διαγράμματος αντιπροσωπεύει μια παρατήρηση. Ακόμη, στα διαγράμματα φυσαλίδας εμφανίζονται οι σχέσεις τουλάχιστον τριών μεταβλητών. Οι δύο από αυτές τις μεταβλητές αποτυπώνονται στον κάθετο και στον οριζόντιο άξονα αντίστοιχα και η τρίτη μεταβλητή αποτυπώνεται στο μέγεθος της εκάστοτε φυσαλίδας.



Σχήμα 4.7: Παράδειγμα Διαγράμματος Φυσαλίδας. (Datavis,2020)

7. Ταμπλό (Dashboard)

Στο χώρο των επιχειρήσεων γίνεται εκτενής χρήση προηγμένων συστημάτων πληροφορικής, τα οποία έχουν την δυνατότητα να επεξεργάζονται γρήγορα, μεγάλες ποσότητες πληροφοριών και να χρησιμοποιούν εξελιγμένες τεχνολογικές πλατφόρμες. Η αξιοποίηση μιας αναβαθμισμένης πλατφόρμας δεν εξυπηρετεί μόνο την ταχύτητα της επεξεργασίας και εξόρυξης πληροφοριών, έχει και τη δυνατότητα συμπύκνωσης και παρουσίασης του μεγάλου αυτού όγκου πληροφορίας με σύγχρονους και αποτελεσματικούς τρόπους.

Με τον όρο ταμπλό ή dashboard, γίνεται αναφορά σε ένα γραφικό περιβάλλον χρήστη, το οποίο εμπεριέχεται σε λογισμικά σύγχρονων συστημάτων πληροφορικής, μέσω του οποίου αποδίδονται δείκτες απόδοσης και προόδου των οργανισμών. Ο τρόπος με τον οποίο αποδίδονται συνοπτικά και παραστατικά τα δεδομένα, θυμίζει το ταμπλό ενός αυτοκινήτου, στο οποίο οι οδηγοί παρακολουθούν, άμεσα, την ορθότητα των λειτουργιών της μηχανής και των ευρύτερων λειτουργικών συστημάτων του οχήματος (Dashboard (business), 2022).

Τα συστήματα Επιχειρηματικής Ευφυΐας χρησιμοποιούν ευρέως τέτοιες μεθόδους επισκόπησης και επιτυγχάνουν, με αυτό τον τρόπο, αύξηση της αποτελεσματικότητάς τους. Τα ηλεκτρονικά λογισμικά προγράμματα στην αγορά, που προσφέρουν δυνατότητα χρήσης ταμπλό ποικίλουν και αναφορικά από τα πιο διάσημα μεταξύ αυτών, είναι:

- Λογισμικό Tableau
- Λογισμικό PowerBi
- Λογισμικό QlikView
- Λογισμικό Google Data View

Στη συνέχεια της ενότητας δίνονται βασικές πληροφορίες για το λογισμικό «Tableau», το οποίο επιλέχθηκε μεταξύ των παραπάνω επιλογών για την εκπόνηση της μελέτης. Βασικά κριτήρια αποτέλεσαν η απλότητα των διεργασιών και το φιλικό περιβάλλοντον διεπαφής χρήστη που παρέχει.

4.3 Λογισμικό Tableau

Το Tableau είναι μία πλατφόρμα οπτικών αναλυτικών στοιχείων που αξιοποιεί με ποικίλες μεθόδους, τον τρόπο με τον οποίο αξιοποιούνται τα δεδομένα. Χρησιμοποιείται ευρέως ως εργαλείο Επιχειρηματικής Ευφυΐας και συμμετέχει ενισχυτικά στην πολυδιάστατη εξερεύνηση και διαχείριση δεδομένων, στην ανακάλυψη και διανομή πληροφοριών σε μικρό χρόνο και υψηλή ταχύτητα. Η χρήση του λογισμικού Tableau δεν απαιτεί υψηλό τεχνικό υπόβαθρο και μεγάλες δεξιότητες προγραμματισμού για να λειτουργήσει. Οι χρήστες προέρχονται από διάφορους επαγγελματικούς κλάδους και εκπαιδευτικές βαθμίδες. Αναλυτές και επιστήμονες δεδομένων, φοιτητές, δάσκαλοι, στελέχη και επιχειρηματίες είναι κάποιες από τις ομάδες, που αξιοποιούν τις δυνατότητες της πλατφόρμας (Tableau, 2022).



Σχήμα 4.8: Tableau Logo (Tableau, 2022)

4.3.1 Η εταιρεία Tableau Software Inc

Η εταιρεία ιδρύθηκε το 2003 στο Mountain View της Καλιφόρνια, διατηρώντας σημερινή έδρα στο Σιάτλ της Ουάσινγκτον. Πρόκειται για μία αμερικανική εταιρεία που εξειδικεύεται κυρίως στη διαδραστική απεικόνιση δεδομένων και ανάπτυξη ευέλικτων λογισμικών που μπορούν να αξιοποιηθούν γύρω από βάσεις δεδομένων, βάσεις δεδομένων τύπου cloud, πίνακες δεδομένων και υπολογιστικά φύλλα.

Βασικοί ιδρυτές της εταιρείας είναι οι Christian Chabot, Pat Hanrahan, Chris Stolte. Τα πρώτα τους βήματα στον κλάδο της διαδραστικής απεικόνισης, έγιναν σε ερευνητικό Τμήμα Επιστήμης Υπολογιστών, με εξειδίκευση στις τεχνικές οπτικοποίησης για ανάλυση σχεσιακών βάσεων και κύβων δεδομένων. Αργότερα, δημιούργησαν ένα κέντρο έρευνας στο Στάνφορντ (1999-2002), το οποίο μετεξελίχθηκε στην δημοφιλή εταιρεία Tableau Software Inc (Tableau Software, 2022).

4.3.2 Προϊόντα της εταιρείας

Για την σωστή επιλογή και χρήση των προϊόντων της εταιρείας Tableau , γίνεται σαφές πως η ανάλυση των δεδομένων μπορεί να ταξινομηθεί σε δύο υποενότητες:

1. Εργαλεία Απαραίτητα για Προγραμματιστές: Η δημιουργία dashboards, γραφημάτων, σύνταξη αναφορών και οπτικοποίησης εμπίπτουν στην εν λόγω κατηγορία. Για να μπορέσει ο χρήστης να εξυπηρετηθεί σε εργασίες σχετικού περιεχομένου με τα παραπάνω, θα χρειαστεί να χρησιμοποιήσει τα προϊόντα Tableau Desktop και Tableau Public.
2. Εργαλεία Κοινής Χρήσης: Στην υποενότητα αυτή, υπάγονται οι διαδικασίες που αφορούν την κοινή χρήση απεικονίσεων, αναφορών και dashboards, που δημιουρ-

γήθηκαν χρησιμοποιώντας εργαλεία προγραμματισμού. Τα προϊόντα της εταιρείας που σχετίζονται με την κοινή χρήση και διανομή των εργασιών ή πληροφοριών είναι τα Tableau Online, Tableau Server και Tableau Reader.

Η κατηγοριοποίηση με βάση το είδος της ανάλυσης των δεδομένων, περιορίζει αντίστοιχα και το εύρος των επιλογών. Στη συνέχεια, δίνονται λεπτομέρειες για κάθε προϊόν ξεχωριστά, προς ανάδειξη του πλέον κατάλληλου, σύμφωνα με τις ανάγκες του χρήστη:

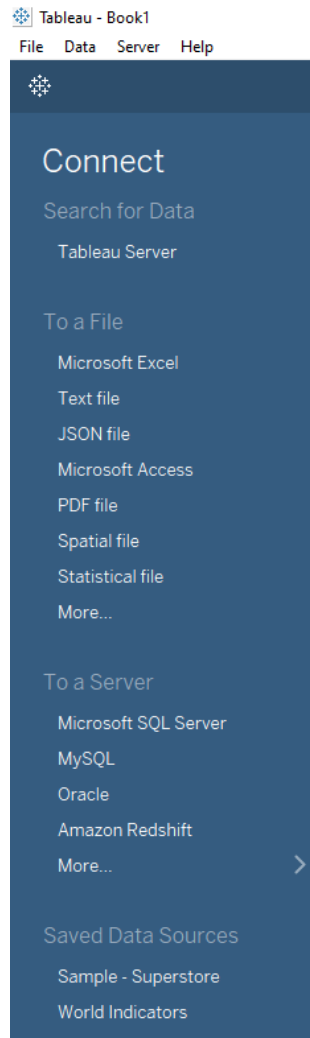
- **Tableau Desktop:** Στο συγκεκριμένο πακέτο παρέχεται η δυνατότητα επεξεργασίας και ανάλυσης με κώδικα, προσαρμογής των αναφορών με βάση τις προτιμήσεις του χρήστη, δημιουργία γραφημάτων μέχρι και συνδυασμό των προηγούμενων με σκοπό την δημιουργία dashboard, για μια πιο γενικευμένη εικόνα των αποτελεσμάτων. Επίσης, το συγκεκριμένο προϊόν δίνει τη δυνατότητα σύνδεσης με αποθήκες δεδομένων, οποιαδήποτε στιγμή, και επιτρέπει την κοινοποίηση των εργασιών που πραγματοποιήθηκαν είτε τοπικά, στο εύρος μιας εταιρείας, είτε δημόσια.
- **Tableau Public:** Βασική διαφορά του συγκεκριμένου προϊόντος από το Tableau Desktop, είναι η αδυναμία αποθήκευσης σε τοπικό επίπεδο. Συγκεκριμένα, σε έναν δημόσιο αποθηκευτικό χώρο, τύπου cloud, της Tableau δημοσιοποιούνται όλες οι εργασίες που προέρχονται από το προϊόν Tableau Public και δεν τις διέπει κανενός είδους απόρρητο ή ασφάλεια προσωπικών στοιχείων. Το Tableau Public συνίσταται σε άτομα που ενδιαφέρονται να γνωρίσουν τις υπηρεσίες του λογισμικού ή όσους θέλουν να κοινοποιήσουν τα δεδομένα τους στο ευρύ κοινό.
- **Tableau Online:** Όπως υποδηλώνει και το όνομα, αφορά ένα διαδικτυακό εργαλείο κοινής χρήσης του Tableau. Τα δεδομένα φιλοξενούνται στον αποθηκευτικό χώρο cloud, που παρέχει η εταιρεία Tableau. Δεν υπάρχει όριο στον όγκο των δεδομένων που μπορούν να δημοσιευτούν στο cloud, όπως επίσης, υπάρχει άμεση

συνδεσιμότητα με περισσότερες απο 40 πηγές δεδομένων, από τις οποίες μπορεί ο χρήστης να συλλέξει δεδομένα, μέσω της πλατφόρμας.

- **Tableau Server:** Οι χρήστες μπορούν να διαμοιράζονται δεδομένα με ασφάλεια που παράγονται από το Tableau Desktop, σε τοπικό επίπεδο, στα πλαίσια της επιχείρησης που το χρησιμοποιεί ή απευθείας σε συστήματα cloud. Τόσο το Tableau Server όσο και το Tableau Online αποσκοπούν στην προβολή των αναλύσεων και την περαιτέρω δημοσιοποίηση τους στο κοινό. Σημειώνεται, πως η επεξεργασία και ανάλυση των δεδομένων, μπορεί να πραγματοποιηθεί μόνο στο περιβάλλον του Tableau Desktop .
- **Tableau Reader:** Αποτελεί δωρεάν εργαλείο, που επιτρέπει την προβολή των εργασιών ή των απεικονίσεων που δημιουργήθηκαν και την δυνατότητα επεξεργασίας των προβολών με χρήση φίλτρων. Οι τροποποιήσεις των αρχείων από τον χρήστη, στο Tableau Reader, είναι αρκετά περιορισμένες.

4.3.3 Βασικές Πληροφορίες Διεπαφής Χρήστη

Το πρώτο βήμα που απαιτείται κατά την έναρξη του προγράμματος, είναι η φόρτωση της πηγής δεδομένων στο σύστημα. Στην εικόνα [4.9], φαίνεται το μενού επιλογών, κατά την εκκίνηση του προγράμματος, με όλους τους πιθανούς τύπους αρχείων που υποστηρίζει το λογισμικό Tableau. Στην παρούσα μελέτη, η βάση δεδομένων βρίσκεται υπό μορφή υπολογιστικού φύλλου Excel. Σε διαφορετική περίπτωση, θα μπορούσαν να χρησιμοποιηθούν αρχεία τύπου Text, PDF, JSON, Microsoft Access, Spatial ή να αξιοποιηθεί κάποιος από τους εγκατεστημένους συνδέσμους που προσφέρει η Tableau, για λήψη βάσης δεδομένων από διακομιστές.



Σχήμα 4.9: Παράθυρο εισόδου δεδομένων στο πρόγραμμα Tableau. (Tableau, 2022)

Με την επιτυχημένη εισαγωγή της επιθυμητής βάσης δεδομένων στο πρόγραμμα, αυτομάτως οδηγούμαστε στο περιβάλλον εργασίας του Tableau. Με εύκολες εναλλαγές μεταξύ δύο καρτελών εργασίας, πραγματοποιείται έλεγχος της βάσης δεδομένων που χρησιμοποιείται, οι μεταβλητές ταξινομούνται σε ποιοτικές και ποσοτικές, συγκρίνονται στοιχεία με απλές τεχνικές «drag and drop» και προκύπτουν αποτελέσματα που μπορούν να παρουσιαστούν είτε μεμονωμένα ως Worksheet, είτε ομαδοποιημένα υπό μορφή dashboard. Στη συνέχεια, παραχωρείται χαρακτηριστική εικόνα του χώρου εργασίας του προγράμματος, στην προσπάθεια περιγραφής των κυριότερων εργαλείων του λογισμικού,

των επιλεγμένων πεδίων δεδομένων και της ευρύτερης πλοήγησης μέσα στο περιβάλλον διεπαφής του χρήστη.



Σχήμα 4.10: Περιβάλλον Εργασίας του προγράμματος Tableau. (Tableau, 2022)

Με βάση το Σχήμα [4.10] και τηρώντας αλφαβητική σειρά, δίνονται περισσότερες πληροφορίες για κάθε επιμέρους καρτέλα ή χωρίο:

Ένδειξη [A]: Εμφανίζει το όνομα της εργασίας, στην οποία αντιστοιχούν φύλλα εργασίας, το πλήθος των οποίων δεν είναι περιορισμένο.

Ένδειξη [B]: «Σέρνοντας» πεδία στις καρτέλες προσθέτουμε και επεξεργαζόμαστε δεδομένα.

Ένδειξη [C]: Πρόκειται για την γραμμή εργασιών. Χρησιμοποιείται για πρόσβαση σε εντολές και εργαλεία ανάλυση ή πλοήγησης.

Ένδειξη [D]: Κύριος χώρος εργασίας. Εδώ οπτικοποιούνται οι διεργασίες.

Ένδειξη [E]: Η επιλογή μας μετακινεί στην κεντρική σελίδα του προγράμματος.

Ένδειξη [F]: Πρόκειται για πλευρικό χωρίο, το οποίο περιέχει τις μεταβλητές της εργασίας, οι οποίες διακρίνονται σε ποιοτικές και ποσοτικές.

Ένδειξη [G]: Η επιλογή μας μετακινεί στη σελίδα Data Source, στην οποία προβάλλονται τα δεδομένα που εισήχθησαν στο Tableau.

Ένδειξη [H]: Εμφανίζει πληροφορίες σχετικά με την τρέχουσα προβολή.

Ένδειξη [I]: Εναλλαγή μεταξύ φύλλων εργασίας.

Κεφάλαιο 5

Μελέτη Περίπτωσης

Ασφαλιστικής Εταιρείας

Στο παρόν κεφάλαιο, αναλύεται η σχέση μίας ασφαλιστικής εταιρείας με το καταναλωτικό της κοινό, χρησιμοποιώντας βάση δεδομένων που διατηρεί η ίδια, για τους πελάτες της. Αρχικά, με χρήση του προγράμματος Excel, γίνεται κατάλληλη επεξεργασία και προετοιμασία του συνόλου των δεδομένων και στη συνέχεια αξιοποιώντας το λογισμικό Tableau, χρησιμοποιούνται μέθοδοι και τεχνικές εξόρυξης πληροφορίας. Τα αποτελέσματα της μελέτης έχουν στόχο, να δώσουν απαντήσεις σε βασικά ερωτήματα, που απασχολούν τη διοίκηση αντίστοιχων εταιρειών και να ενισχύσουν περαιτέρω τη διαδικασία λήψης αποφάσεων. Σημειώνεται, πως η συλλογή, αποθήκευση και επεξεργασία των δεδομένων της βάσης των πελατών της ασφαλιστικής εταιρείας, τηρεί την προστασία δεδομένων στο πλαίσιο του ΓΚΠΔ (Γενικός Κανονισμός Προστασίας Δεδομένων), κάνοντας χρήση δεδομένων, διατηρώντας όμως πάντα πλήρη ανωνυμία.

5.1 Βάση Δεδομένων της Ασφαλιστικής Εταιρείας

Η βάση δεδομένων, στην οποία βασίστηκε το μεγαλύτερο μέρος της συγκεκριμένης μελέτης, βρίσκεται σε μορφή πίνακα. Κάθε σειρά του πίνακα αντιπροσωπεύει έναν μοναδικό πελάτη της ασφαλιστικής εταιρείας και κάθε στήλη του πίνακα αντιπροσωπεύει μία ιδιότητα ή χαρακτηριστικό που αποδίδεται σε κάθε πελάτη. Συγκεκριμένα, ο αριθμός των καταχωρήσεων (στήλες του πίνακα) ή ο συνολικός αριθμός πελατών, που εμπεριέχονται στη βάση δεδομένων, ανέρχεται στους 10.008. Ο αριθμός των στηλών της βάσης δεδομένων ανέρχεται στις 15 και στη συνέχεια δίνονται πληροφορίες για κάθε στήλη του πίνακα, ξεχωριστά:

1. **Group Code:** Στην στήλη καταχωρείται το ασφαλιστικό πακέτο που κάθε πελάτης έχει λάβει. Οι τιμές της μεταβλήτης Group Code είναι αποκλειστικά αριθμοί και σε κάθε ασφαλιστικό πακέτο δίνεται μια μοναδική κωδικοποίηση.

- Ασφάλεια Ζωής: «1»
- Ασφάλεια Υγείας: «4»
- Ασφάλεια Σπιτιού: «17»
- Ασφάλεια Αυτοκινήτου: «19»
- Λοιπές Ασφάλειες: «3», «15», «16», «18», «22»

2. **Group Code Dimension:** Πρόκειται για παρόμοια στήλη με αυτή της Group Code με την μόνη διαφορά, πως τα στοιχεία δεν είναι αριθμητικά αλλά χαράκτῆρες(λέξεις).

- Ασφάλεια Ζωής: «LIFE»
- Ασφάλεια Υγείας: «HEALTH»
- Ασφάλεια Σπιτιού: «HOUSE»

- Ασφάλεια Αυτοκινήτου: «CAR»
 - Λοιπές Ασφαλίσεις: «REST»
3. **Status Description:** Εμφανίζει την κατάσταση του συμβολαίου κάθε πελάτη, μεταξύ των επιλογών «ΑΚΥΡΟ» ή «ΕΝΕΡΓΟ».
4. **Status Code:** Πρόκειται για παρόμοια στήλη με την προηγούμενη, με την μόνη διαφορά πως οι επιλογές «ΑΚΥΡΟ» και «ΕΝΕΡΓΟ» , έχουν αντικατασταθεί απο τις τιμές «0» και «2» αντίστοιχα.
5. **Customer Insert Date:** Εδώ καταγράφονται οι ημερομηνίες έναρξης ασφάλισης κάθε πελάτη.
6. **Customer Cancel Date:** Εδώ καταγράφονται οι ημερομηνίες ακύρωσης ασφάλισης κάθε πελάτη.
7. **Customer Code:** Πρόκειται για μια μοναδική ψηφιακή ταυτότητα, που δίνεται από την εταιρεία στον κάθε πελάτη.
8. **Portal:** Πρόκειται για μία ψηφιακή πλατφόρμα διαχείρισης, στην οποία οι χρήστες μπορούν να ελέγχουν και να ενημερώνονται για θέματα που αφορούν τα ασφαλιστικά τους προϊόντα, τη δυνατότητα πλήρωμής των συμβολαίων τους, τη μετατροπή ή την ακύρωση κάποιου συμβολαίου.
9. **Channel Class:** Η μεταβλητή αφορά το δίκτυο των ασφαλιστών. Λαμβάνει τις εξής τιμές:
- Δίκτυο Agency: Δίκτυο της ασφαλιστικής εταιρείας που προωθεί μέσω ασφαλιστών, αποκλειστικά προϊόντα και υπηρεσίες της συγκεκριμένης εταιρείας.
 - Πρακτορειακό Δίκτυο: Δίκτυο ασφαλιστικών συμβούλων μη αποκλειστικής συνεργασίας. Πρόκειται για ασφαλιστικά γραφεία, που συγκεντρώνουν με-

μονωμένους ασφαλιστές και προωθούν στους πελάτες την πιο συμφέρουσα και προτιμητέα προσφορά ασφάλισης, μεταξύ πολλών και διαφορετικών ασφαλιστικών εταιρειών.

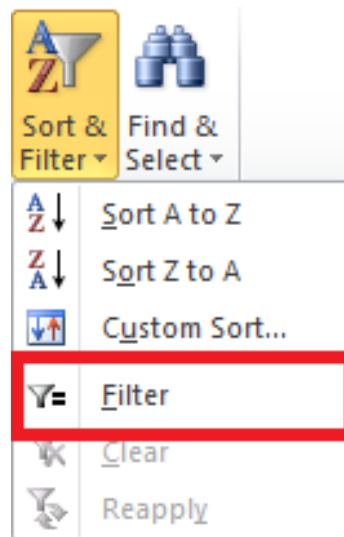
- Μέριμνα: Πρόκειται για θυγατρική της ασφαλιστικής εταιρείας, που λειτουργεί ως ασφαλιστικό γραφείο, μη αποκλειστικής συνεργασίας.
- Τραπεζικό Δίκτυο: Συνεργασία της ασφαλιστικής εταιρείας με τράπεζα για την προώθηση ασφαλιστικών προϊόντων.
- Ίδιο Δίκτυο: Αποτελείται από ιδιώτες πελάτες που ασφαλίζονται απευθείας με την εταιρεία, χωρίς τη μεσολάβηση ασφαλιστικού συμβούλου.

10. **Payment Frequency**: Συχνότητα πληρωμής του ασφαλιστικού συμβολαίου. Οι πιθανές τιμές της μεταβλητής είναι «ΤΡΙΜΗΝΗ», «ΕΞΑΜΗΝΗ» και «ΕΤΗΣΙΑ» και «ΕΦΑΠΑΞ».
11. **Num of Payment Annually**: Περιγράφει τη συχνότητα πληρωμής του πελάτη, με τη μόνη διαφορά πως οι τιμές των μεταβλητών για τρίμηνη, εξάμηνη και ετήσια συχνότητα είναι «4», «2» και «1» αντίστοιχα.
12. **Gender**: Πρόκειται για το βιολογικό φύλο κάθε πελάτη και οι τιμές που αποδίδονται στην μεταβλητή «MALE», «FEMALE» και «OTHER». Η τελευταία τιμή αφορά νομική οντότητα και όχι κάποιο φυσικό πρόσωπο.
13. **Customer Occupation Name**: Η επαγγελματική απασχόληση κάθε πελάτη.
14. **Customer Age**: Η ηλικία κάθε πελάτη
15. **Geo Area**: Η γεωγραφική θέση (περιφέρεια) που υπάγεται κάθε πελάτης, με πιθανές τιμές για την μεταβλητή: «ΑΤΤΙΚΑ», «ΜΑΚΕΔΟΝΙΑ», «ΠΕΛΟΠΟΝΝΗΣΗ», «ΚΡΗΤΗ», «REST».

5.2 Επεξεργασία και Προετοιμασία Βάσης Δεδομένων

Η διαδικασία επεξεργασίας και προετοιμασίας της βάσης δεδομένων χωρίστηκε σε βήματα. Βασικός σκοπός των ενεργειών που εκτελέστηκαν ήταν κυρίως η διόρθωση ορθογραφικών λαθών, η μετατροπή πεζών γραμμάτων σε κεφαλαία με σκοπό την επίτευξη ομοιομορφίας των τιμών, συμπλήρωση κενών στοιχείων και διόρθωση των αποστάσεων μεταξύ χαρακτήρων.

Βήμα 1ο: Ξεκινώντας την διαδικασία επεξεργασίας και διόρθωσης, χρησιμοποιήσαμε το εργαλείο «Sort & Filter» του υπολογιστικού φύλλου Excel. Στην οθόνη εμφανίζεται το χαρακτηριστικό μενού φίλτρων και διάταξης και επιλέγουμε την ένδειξη «Filter», [Σχήμα 5.1].



Σχήμα 5.1: Εργαλείο Ταξινόμησης και Φίλτρων στο Excel

Με τον τρόπο αυτό, δίνεται η δυνατότητα ταξινόμησης και φιλτραρίσματος κάθε στήλης ξεχωριστά και μπορούν να εντοπιστούν τα κενά στοιχεία καθώς και οποιαδήποτε ανομοιογένεια, γρήγορα και αποτελεσματικά. Η μπάρα των στηλών μετά την επιλογή του Sort & Filter θα πρέπει να έχει τη μορφή όπως στο Σχήμα 5.2.

	D	E	F	G	H	I	J
1	STATUS_DESCR	CANCEL_DATE	CUSTOMER_INSERT_DATE	CUSTOMER_CODE	PORTAL	CHANNEL_CLASS	PAYMENT_FREQUENCY

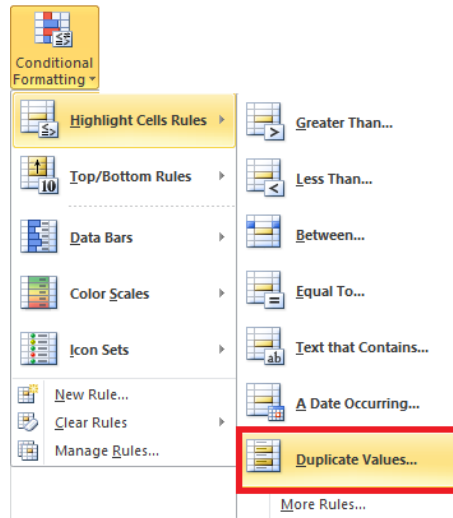
Σχήμα 5.2: Στήλες της Βάσης Δεδομένων μετά την ενεργοποίηση του εργαλείου Sort & Filter

Βήμα 2ο: Αφού εφαρμόστηκαν φίλτρα στις στήλες της βάσης δεδομένων, έγιναν οι απαραίτητες ορθογραφικές διορθώσεις σε κάθε στήλη ξεχωριστά, εντοπίστηκαν και αφαιρέθηκαν ακραίες τιμές, συμπληρώθηκαν κενά στοιχεία και πεζά γράμματα μετατράπηκαν σε κεφαλαία. Επίσης, δόθηκε έμφαση στην στήλη «CUSTOMER OCCUPATION NAME», όπου παρατηρήθηκε μεγάλη ποικιλία επαγγελμάτων, η οποία δεν διευκόλυνε τη διαδικασία ομαδοποίησης και ανάδειξης συγκεκριμένων target group. Για τον λόγο αυτό, τα επαγγέλματα συγχωνεύτηκαν σε μεγαλύτερες κατηγορίες και οι διαθέσιμες επιλογές, είναι: «ΣΠΟΥΔΑΣΤΗΣ», «ΑΝΕΡΓΟΣ», «ΜΙΣΘΩΤΟΣ», «ΕΛΕΥΘΕΡΟΣ ΕΠΑΓΓΕΛΜΑΤΙΑΣ», «ΣΥΝΤΑΞΙΟΤΧΟΣ», «ΟΙΚΟΚΥΡΙΚΑ».

Βήμα 3ο: Στο τελευταίο βήμα προετοιμασίας της λίστας, εφαρμόσαμε την μέθοδο εύρεσης διπλοτιμών στην κατηγορία «Customer Code». Η χρήση αυτού του εργαλείου στα υπολογιστικά φύλλα του Excel, επισημαίνει με έντονο χρώμα τις καταχωρήσεις που φέρουν ίδια τιμή. Συνεπώς, κάθε ιδιοτιμή στην κατηγορία αυτή, υποδυκνύει πως ο ίδιος πελάτης έχει παραπάνω από ένα ασφαλιστικό προϊόν στην εταιρεία.

5.3 Εξόρυξη Πληροφορίας Μέσω του Λογισμικού Tableau

Με το πέρας της διαδικασίας οργάνωσης και τακτοποίησης των δεδομένων, γίνεται χρήση του λογισμικού Tableau, για εξόρυξη πληροφορίας και οπτικοποίηση των αποτελεσμάτων. Στην συγκεκριμένη μελέτη, αξιοποιήθηκε το προϊόν «Tableau Desktop», το οποίο παρέχεται δωρεάν, για περιορισμένο χρονικό διάστημα σε σπουδαστές.



Σχήμα 5.3: Εργαλείο ιδιοτιμών σε φύλλο Excel

Με βασικά κριτήρια την πρακτική ανάλυση των δεδομένων και την ανάθεση εύστοχων επιχειρηματικών στόχων στην εταιρεία, επιχειρείται στην παρούσα υποενότητα η διεξαγωγή υποθετικών ερωτημάτων, τα οποία στη συνέχεια θα απαντηθούν κάνοντας χρήση οπτικοποιήσεων του Tableau.

Ερώτημα 1: Με γνώμονα τα Ασφαλιστικά Προϊόντα, που παρέχει η ασφαλιστική εταιρεία, τι συμπεράσματα προκύπτουν σχετικά με τα Δίκτυα Ασφάλισης της εταιρείας και τους πελάτες της;

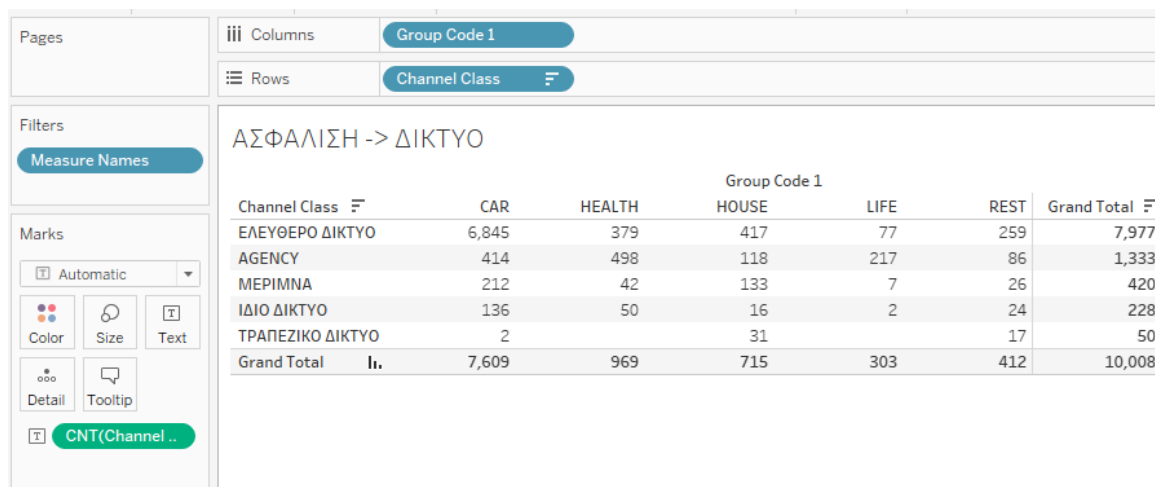
Για να δοθεί απάντηση στο πρώτο ερώτημα, αξιοποιήθηκαν δεδομένα της βάσης, που αφορούν:

1. τα Ασφαλιστικά Προϊόντα της εταιρείας, δηλαδή χρησιμοποιήθηκε το πεδίο της βάσης δεδομένων «**Group Code**».
2. τα Δίκτυα Ασφάλισης της εταιρείας, δηλαδή χρησιμοποιήθηκε το πεδίο της βάσης δεδομένων «**Channel Class**».

3. τα στοιχεία που αφορούν το καταναλωτικό κοινό της εταιρείας, δηλαδή χρησιμοποιήθηκαν τα πεδία της βάσης «**Gender**», «**Geo Area**» και «**Customer Occupation Name**».

Στη συνέχεια, συσχετίζεται η μεταβλητή «Group Code» με κάθε μία από τις υπόλοιπες μεταβλητές, στο περιβάλλον εργασίας του Tableau. Κάνοντας χρήση «Drag and Drop» κινήσεων, τοποθετείται κάθε φορά στο χωρίο Columns η μεταβλητή «Group Code» και για κάθε διαφορετική περίπτωση συσχέτισης, τοποθετείται στο χωρίο Rows η ανάλογη μεταβλητή.

Στο Σχήμα [5.4] που ακολουθεί, φαίνονται τα αποτελέσματα της πρώτης συσχέτισης μεταξύ Ασφαλιστικών Προϊόντων της εταιρείας και Δικτύων Ασφάλισης.



The screenshot shows the Tableau interface with the following components:

- Columns:** Group Code 1
- Rows:** Channel Class
- Filters:** Measure Names
- Marks:** Automatic
- Table Title:** ΑΣΦΑΛΙΣΗ -> ΔΙΚΤΥΟ

Channel Class	CAR	HEALTH	HOUSE	LIFE	REST	Grand Total
ΕΛΕΥΘΕΡΟ ΔΙΚΤΥΟ	6,845	379	417	77	259	7,977
AGENCY	414	498	118	217	86	1,333
ΜΕΡΙΜΝΑ	212	42	133	7	26	420
ΙΔΙΟ ΔΙΚΤΥΟ	136	50	16	2	24	228
ΤΡΑΠΕΖΙΚΟ ΔΙΚΤΥΟ	2		31		17	50
Grand Total	7,609	969	715	303	412	10,008

Σχήμα 5.4: Συσχέτιση Μεταβλητών Ασφαλιστικά Προϊόντα - Δίκτυα Ασφάλισης

Όπως προαναφέρθηκε, το χωρίο Columns (στήλες), περιλαμβάνει τα Προϊόντα Ασφάλισης ενώ το χωρίο Rows (Γραμμές), περιλαμβάνει τα Δίκτυα Ασφάλισης. Στην αριστερή πλευρά, στο Σχήμα[5.4], παρατηρείται και το εργαλείο «Filters» με το οποίο μπορεί να εφαρμοστεί φίλτρο, που απομονώνει κάποιο συγκεκριμένο στοιχείο ή απορρίπτει στοιχεία του πίνακα που δεν έχουμε ανάγκη. Με αυτό τον τρόπο η διαδικασία οπτικοποίησης συγκεκριμενοποιείται ευκολότερα και μειώνεται ο χρόνος ανάλυσης.

Παρατηρήσεις:

- Έχει οριστεί φθίνουσα σειρά εμφάνισης των αποτελεσμάτων με βάση την τελευταία στήλη του πίνακα, στην οποία παρουσιάζονται το σύνολο των πελατών που ανήκουν σε κάθε Ασφαλιστικό Δίκτυο.
- Στην τελευταία στήλη του πίνακα εμφανίζεται το σύνολο των πελατών της εταιρείας για κάθε επιμέρους ασφαλιστικό προϊόν.
- Η μεγαλύτερη μερίδα του καταναλωτικού κοινού της εταιρείας εντοπίζεται στο ασφαλιστικό προϊόν «CAR» και το δίκτυο με τη μεγαλύτερη πελατιακή βάση είναι το Ελεύθερο Δίκτυο Ασφαλιστών.
- Το στοιχείο του πίνακα με Ασφαλιστικό Προϊόν «**CAR**» και με Δίκτυο Ασφάλισης «**ΕΛΕΥΘΕΡΟ ΔΙΚΤΥΟ**» είναι το πιο πολυπληθές.

Ακολουθούν, κατά παρόμοιο τρόπο, τα αποτελέσματα των συσχετίσεων μεταξύ Ασφαλιστικών προϊόντων της εταιρείας και των «Gender», «Geo Area» και «Customer Occupation».

Pages

Filters

Marks

Automatic

Color

Size

Text

Detail

Tooltip

CNT(Gender)

Columns

Group Code 1

Rows

Gender

ΑΣΦΑΛΙΣΗ -> ΦΥΛΟ

	Group Code 1					
Gender	CAR	HEALTH	HOUSE	LIFE	REST	Grand Total
MALE	4,754	506	337	140	125	5,862
FEMALE	2,610	375	255	122	40	3,402
OTHER	245	88	123	41	247	744

Σχήμα 5.5: Συσχέτιση Μεταβλητών: Ασφαλιστικά Προϊόντα - Βιολογικό Φύλο Πελάτη

Στο Σχήμα [5.5], φαίνονται τα αποτελέσματα της δεύτερης συσχέτισης μεταξύ Ασφαλιστικών Προϊόντων της εταιρείας και του Βιολογικού Φύλου κάθε πελάτη.

Παρατηρήσεις:

- Το μεγαλύτερο πελατειακό κοινό της εταιρείας, σύμφωνα με τα δεδομένα της βάσης δεδομένων, είναι ανδρικού γένους. Συγκεκριμένα, ο αριθμός πελατών ανδρικού γένους ανέρχεται στους 5.862, ενώ ο αριθμός πελατών θηλυκού γένους ανέρχεται στις 3,402.
- Οι υπόλοιπες 744 περιπτώσεις είναι καταχωρημένες ως «OTHER» και πρόκειται για μη-φυσικά πρόσωπα, δηλαδή νομικές οντότητες στο χώρο των επιχειρήσεων.
- Το στοιχείο του πίνακα με Ασφαλιστικό Προϊόν «**CAR**» στη γεωγραφική περιφέρεια «**ΑΤΤΙΚΑ**», είναι το πιο πολυπληθές.

Pages

Filters

Marks

Columns

Rows

Geo Area

Automatic

Color

Size

Text

Detail

Tooltip

CNT(Geo Area)

ΑΣΦΑΛΙΣΗ -> ΓΕΩΓ. ΠΕΡΙΟΧΗ

	Group Code 1					
Geo Area	CAR	HEALTH	HOUSE	LIFE	REST	Grand Total
ΑΤΤΙΚΑ	4,435	696	329	194	120	5,774
MACEDONIA	1,357	71	75	13	35	1,551
CRETE	1,271	20	50	6	15	1,362
REST	357	66	106	46	42	617
PELOPONNESE	189	17	39	19	10	274

Σχήμα 5.6: Συσχέτιση Μεταβλητών: Ασφαλιστικά Προϊόντα - Γεωγραφική Περιφέρεια Πελάτη

Στο Σχήμα [5.6], φαίνονται τα αποτελέσματα της τρίτης συσχέτισης μεταξύ Ασφαλιστικών Προϊόντων της εταιρείας και της Γεωγραφικής Περιοχής που ανήκει κάθε πελάτη.

Παρατηρήσεις:

- Η Αττική αποτελεί την γεωγραφική περιφέρεια με την μεγαλύτερη πελατεία της εταιρείας. Ακολουθούν η Μακεδονία και η Κρήτη, με σχεδόν ισόποσα σύνολα πελατών.
- Η τελευταία στήλη του πίνακα εμφανίζει το σύνολο των πελατών της εταιρείας, για κάθε επιμέρους γεωγραφική περιφέρεια και ακολουθεί φθίνουσα σειρά αποτελεσμάτων.
- Το στοιχείο του πίνακα Ασφαλιστικό Προϊόν «**CAR**» και επιλογή Βιολογικού Φύλου «**MALE**», είναι το πιο πολυπληθές.

Pages

Filters

Marks

Automatic

Color

Size

Text

Detail

Tooltip

CNT(Custome..

Columns

Group Code 1

Rows

Customer Occupa..

ΑΣΦΑΛΙΣΗ -> ΕΠΑΓΓ. ΑΠΑΣΧΟΛΗΣΗ

Group Code 1

Customer Occupatio..	CAR	HEALTH	LIFE	HOUSE	REST	Grand Total
ΜΙΣΘΩΤΟΣ	5,106	470	95	311	47	6,029
ΕΛΕΥΘΕΡΟΣ ΕΠΑΓΓΕΛΜΑ..	1,439	354	150	298	344	2,585
ΣΥΝΤΑΞΙΟΥΧΟΣ	751	54	23	70	13	911
ΟΙΚΟΚΥΡΙΑ	120	37	23	21	3	204
ΑΝΕΡΓΟΣ	111	33	4	3	4	155
ΣΠΟΥΔΑΣΤΗΣ	82	21	8	12	1	124

Σχήμα 5.7: Συσχέτιση Μεταβλητών: Ασφαλιστικά Προϊόντα - Επαγγελματική Απασχόληση Πελάτη

Στο Σχήμα [5.7], φαίνονται τα αποτελέσματα της τέταρτης συσχέτισης μεταξύ Ασφαλιστικών Προϊόντων της εταιρείας και της Επαγγελματικής Απασχόλησης κάθε πελάτη.

Παρατηρήσεις:

- Η πολυπληθέστερη τιμή στην μεταβλήτη «Customer Occupation Name» είναι η τιμή «ΜΙΣΘΩΤΟΣ», με σύνολο 6.029 τιμές.
- Η τελευταία στήλη του πίνακα εμφανίζει το σύνολο των πελατών της εταιρείας, για κάθε επιμέρους επαγγελματική απασχόληση και ακολουθεί φθίνουσα σειρά αποτελεσμάτων.
- Το στοιχείο του πίνακα με Ασφαλιστικό Προϊόν «**CAR**» και επιλογή Ασφαλιστικής Απασχόλησης «**ΜΙΣΘΩΤΟΣ**», είναι το πιο πολυπληθές.

Ερώτημα 2: Ελέγχοντας το χρόνο παραμονής -«αφοσίωση»- των πελατών στην εταιρεία, πώς σχετίζεται το κριτήριο Συχνότητας Πληρωμών και ποιά από τις διαθέσιμες επιλογές θα πρέπει να παρέχεται κατά κύριο λόγο από την εταιρεία;

Για να δοθεί απάντηση στο δεύτερο ερώτημα, αξιοποιήθηκαν δεδομένα της βάσης, που αφορούν:

1. τη Συχνότητα Πληρωμής του ασφαλιστικού συμβολαίου από τους πελάτες, δηλαδή χρησιμοποιήθηκε το πεδίο της βάσης δεδομένων «**Payment Frequency**».
2. τα Δίκτυα Ασφάλισης της εταιρείας, δηλαδή χρησιμοποιήθηκε το πεδίο της βάσης δεδομένων «**Channel Class**».
3. η ημερομηνία έναρξης ασφάλισης του κάθε πελάτη, δηλαδή χρησιμοποιήθηκε το πεδίο της βάσης δεδομένων «**Customer Insert Date**».

Στη συνέχεια, μέσω της διαδικασίας συσχέτισης της κύριας μεταβλητής του ερωτήματος «**Payment Frequency**» προκύπτει το παρακάτω ταμπλό (dashboard) αποτελεσμάτων:

Συχνότητα Πληρωμής / Ημερομηνία Έναρξης Συμβολαίου

Year of Cus..	Payment Frequency			
	ΕΞΑΜΗΝΟΣ	ΕΤΗΣΙΟΣ	ΤΡΙΜΗΝΟΣ	ΕΦΑΠΑΞ
2000	4	1		
2001	509	196	58	14
2002	178	36	22	
2003	135	47	17	
2004	129	69	9	2
2005	87	69	8	1
2006	588	384	84	21
2007	74	54	9	1
2008	82	43	4	2
2009	215	151	15	2
2010	667	122	38	5
2011	445	134	34	5
2012	608	143	20	6
2013	635	185	50	10
2014	781	116	60	3
2015	678	104	65	4
2016	693	117	123	5
2017	506	119	199	7
2018		1		
Grand T.. II.	7,014	2,091	815	88

Συχνότητα Πληρωμής / Δίκτυο

Channel Class	Payment Frequency				Grand Total
	ΕΞΑΜΗΝΟΣ	ΕΤΗΣΙΟΣ	ΕΦΑΠΑΞ	ΤΡΙΜΗΝΟΣ	
AGENCY	611	405	42	275	1,333
ΕΛΕΥΘΕΡΟ ΔΙΚΤΥΟ	6,046	1,365	40	526	7,977
ΙΔΙΟ ΔΙΚΤΥΟ	139	84	1	4	228
ΜΕΡΙΜΝΑ	200	205	5	10	420
ΤΡΑΠΕΖΙΚΟ ΔΙΚΤΥΟ	18	32			50

Σχήμα 5.8: Dashboard Ερωτήματος 2

Παρατηρήσεις:

- Περίπου το 70% των καταχωρήσεων στη μεταβλητή «Payment Frequency» έχει λάβει την τιμή «ΕΞΑΜΗΝΟΣ», συνεπώς η συντριπτική πλειοψηφία των πελατών της ασφαλιστικής εταιρείας επιλέγουν πληρωμή του συμβολαίου τους κάθε εξάμηνο. Ακολουθεί η επιλογή πληρωμής «ΕΤΗΣΙΑ» και ύστερα οι τιμές «ΤΡΙΜΗΝΟΣ» και «ΕΦΑΠΑΞ».

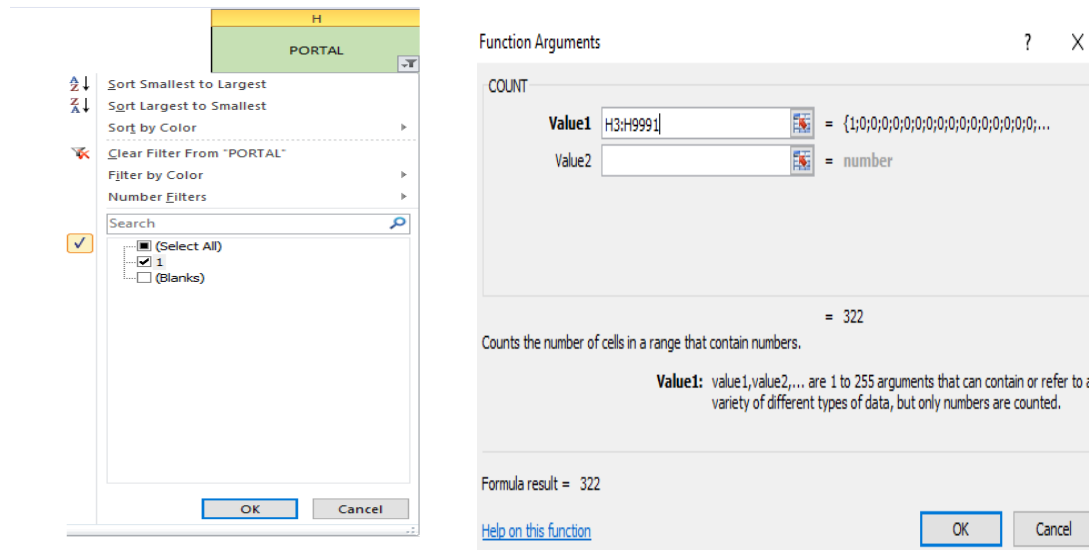
- Η τελευταία σειρά του πίνακα, εμφανίζει το σύνολο των πελατών της εταιρείας, για κάθε επιμέρους συχνότητα πληρωμής ασφαλιστικού συμβολαίου, στην οποία εφαρμόστηκε φθίνουσα σειρά αποτελεσμάτων.

Από τα αποτελέσματα της ανάλυσης φαίνεται πως, βασική παρότρυνση της εταιρείας προς τους καταναλωτές της είναι η αποπληρωμή συμβολαίου ανά εξάμηνο ή με το πέρας κάθε έτους. Με αυτή την τακτική, μπορούν να γνωρίζουν ότι ο πελάτης παραμένει για μεγάλο διάστημα του τρέχοντος έτους «πιστός» στην εταιρεία και παράλληλα δεν θα ενοχλείται με υπενθυμίσεις για τακτικές πληρωμές, που θα έπρεπε να πραγματοποιεί σε περιπτώσεις μηνιαίας ή τρίμηνης συχνότητας πληρωμών.

Ερώτημα 3: Με ποió τρόπο έχει συνεισφέρει η Διαδικτυακή Πύλη (Portal) της εταιρείας;

Η προσπάθεια ανάλυσης των δεδομένων που αφορούν τη Διαδικτυακή Πύλη όσο και την εύρεση απάντησης στο Ερώτημα 3, μας οδηγεί στην χρήση των υπολογιστικών εργαλείων του προγράμματος Excel και ύστερα στη δημιουργία οπτικοποίησης μέσω του λογισμικού Tableau.

Αρχικά, κάνοντας χρήση των φίλτρων [5.1], που εισήχθησαν στο φύλλο Excel προηγουμένως, βρίσκουμε τις καταχωρήσεις που χρησιμοποιούν Portal και μετράμε το πλήθος τους. Σημειώνεται, πως οι περιπτώσεις πελατών που χρησιμοποιούν τη Διαδικτυακή Πύλη φέρουν την τιμή «1» στη μεταβλητή Portal, ενώ οι περιπτώσεις που δεν χρησιμοποιούν τη Διαδικτυακή Πύλη δεν φέρουν κάποιο αριθμό ή χαρακτήρα (Blank). Για την καταμέτρηση χρησιμοποιήθηκε η εντολή COUNT στο υπολογιστικό φύλλο του Excel, η οποία μετράει μόνο το πλήθος των κελιών με αριθμητικές τιμές και αγνοεί κελιά που περιέχουν χαρακτήρες ή κενά.



(α') Επιλογή φίλτρου στην κατηγορία Portal

(β') Ορίσματα συνάρτησης COUNT

Σχήμα 5.9

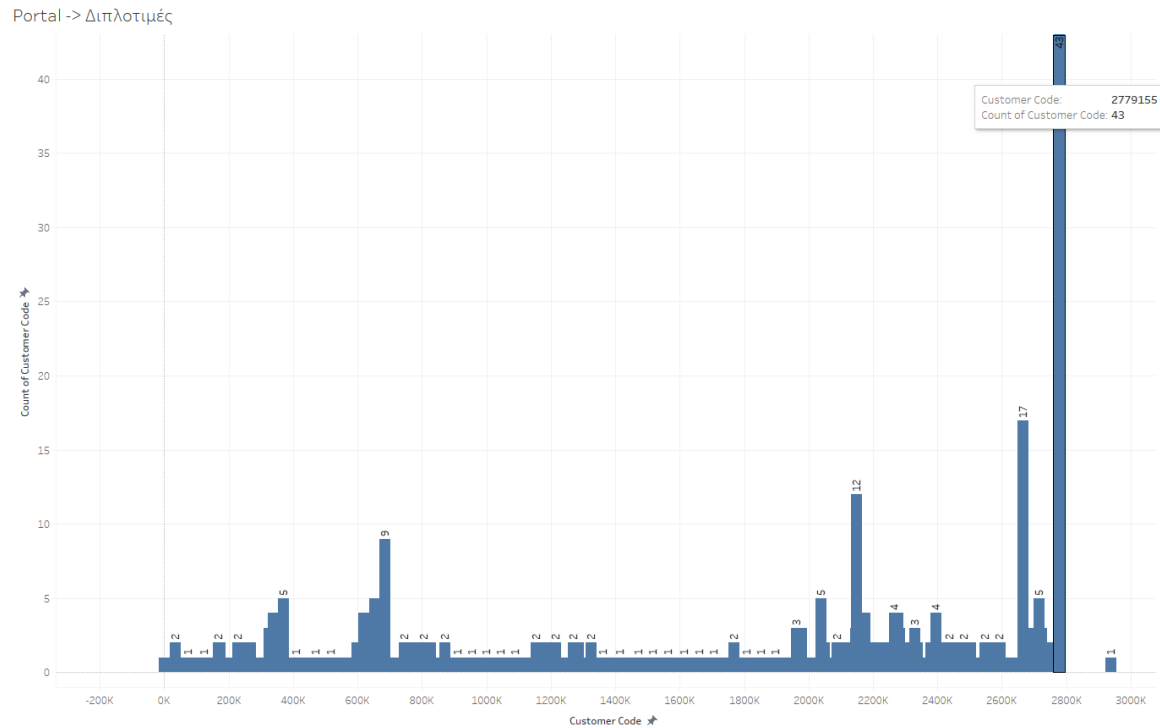
Όπως φαίνεται και στο παράθυρο ορισμάτων της συνάρτησης COUNT Σχήμα[5.9], ο συνολικός αριθμός των καταχωρήσεων που χρησιμοποιούν Διαδικτυακή Πύλη, είναι 322.

Στη συνέχεια, αξιοποιήθηκαν δεδομένα της βάσης, που αφορούν:

1. τον μοναδικό κώδικό κάθε καταχώρησης, δηλαδή χρησιμοποιήθηκε το πεδίο της βάσης δεδομένων «**Customer Code**».
2. τους πελάτες που χρησιμοποιούν Διαδικτυακή Πύλη, δηλαδή χρησιμοποιήθηκε το πεδίο της βάσης δεδομένων «**Portal**».

Ζητείται από το Tableau, να οπτικοποιήσει το σύνολο των πελατών που χρησιμοποιούν την εφαρμογή Portal και τα αποτελέσματα φαίνονται στο παρακάτω Σχήμα.

Κάθε ράβδος του ραβδογράμματος εκπροσωπεί έναν πελάτη της εταιρείας, που χρησιμοποιεί τη Διαδικτυακή Πύλη. Οι αριθμοί που αναγράφονται στο πάνω μέρος κάθε ράβδου είναι ο αριθμός των ασφαλιστικών προϊόντων που διατηρεί με την εταιρεία ο πελάτης.



Σχήμα 5.10: Οπτικοποίηση Πελατών που χρησιμοποιούν Portal

Παρατηρήσεις:

- Αρκετοί από τους χρήστες Portal, έχουν παραπάνω του ενός προϊόντος ασφάλισης. Συνεπώς, παρά την μικρή δημοτικότητα της Διαδικτυακής Πύλης στο δείγμα, φαίνεται πως η προσπάθειες περαιτέρω προώθησής της στους πελάτες, θα τους οδηγήσει σε απόκτηση επιπλέον προϊόντων ασφάλισης. Με αυτό τον τρόπο το καταναλωτικό κοινό της εταιρείας μπορεί να έχει πλήρη έλεγχο και ενημέρωση σχετικά με τα ασφαλιστικά του προϊόντα και συγχρόνως η εταιρία αυξάνει αποτελεσματικά την κερδοφορία της.

Ερώτημα 4: Ποιά ηλικιακή ομάδα πελατών έχει τη μεγαλύτερη ασφαλιστική συνείδηση;

Η απάντηση στο ερώτημα απαιτεί την αξιοποίηση των εξής δεδομένων:

1. Ηλικίες του κάθε πελάτη της εταιρείας, δηλαδή χρησιμοποιείται το πεδίο της βάσης δεδομένων «**Customer Age**»[5.1].

Το Tableau δίνει την δυνατότητα συσταδοποίησης των δεδομένων, κάνοντας χρήση του αλγόριθμου K-means. Τα αποτελέσματα της μεθόδου φαίνονται στο παρακάτω Σχήμα[5.11].



Σχήμα 5.11: Αλγόριθμος K-means

Παρατηρήσεις:

- Για την διαδικασία συσταδοποίησης τοποθετήθηκαν στον οριζόντιο άξονα όλες οι ηλικίες των καταχωρημένων πελατών της εταιρείας. Στη συνέχεια με την εντολή count στην μεταβλητή «**Customer Age**» ζητήθηκε στο πρόγραμμα να καταχωρήσει σε κάθε ηλικία το πλήθος των πελατών που είχαν την ηλικία αυτή.
- Τέλος, με απλή χρήση Drag and Drop εφαρμόστηκε η μέθοδος K-means επιλέγοντας προκαθορισμένη ποσότητα κεντροειδών $k=3$.
- Τα αποτελέσματα του Σχήματος [5.11] παρουσιάζουν τρεις ηλικιακές συστάδες. Η **πρώτη** συστάδα κυμαίνεται [20,24]U[74,99] (μπλε χρώμα), η **δεύτερη** συστάδα κυμαίνεται [25,35]U[61,73] (κίτρινο χρώμα) και η **τρίτη** συστάδα κυμαίνεται [36,60] (κόκκινο χρώμα).

Το μεγαλύτερο σύνολο καταναλωτών της εταιρείας βρίσκεται στην τρίτη ηλικιακή συστάδα με ηλικίες 36 έως 60. Στον αντίποδα, μικρότερη ασφαλιστική συνείδηση εντοπίζεται στο πρώτο ηλικιακό γκρουπ που αποτελείται από τους νεότερους και του γηραιότερους πελάτες της βάσης δεδομένων. Τέλος, φαίνεται πως ο διαχωρισμός των συστάδων, με αρχικό πλήθος $k=3$, οδήγησε σε ικανοποιητικά αποτελέσματα με ξεκάθαρο σχηματισμό αυτόνομων συστάδων και δεν χρειάστηκε περαιτέρω επεξεργασία και δοκιμές του αλγορίθμου.

Σημαντικό Γνωστικό Υπόβαθρο:

- Η υποχρεωτική ασφάλιση αυτοκινήτου, με βάση το νόμο, είναι ο βασικός λόγος για τον οποίο, το πλήθος των πελατών με ασφάλιση αυτοκινήτου υπερτερεί αισθητά έναντι των υπόλοιπων ασφαλίσεων της βάσης.
- Η δράση της εταιρείας, έχει επικεντρωθεί στην περιοχή της Αθήνας. Αυτό αποτελεί στρατηγική που έχει λάβει, εν γνώση της, η εταιρεία, καθώς φέρει συγκεντρωμένα και το μεγαλύτερο ποσοστό πληθυσμού της χώρας.

- Με τον όρο χαμηλή ασφαλιστική συνείδηση εννοείται η προτίμηση του καταναλωτικού κοινού να προτιμά μόνο την ασφάλιση αυτοκινήτου και να μην μεριμνεί περαιτέρω για πιθανές ζημιές ή ατυχήματα σε άλλο κλάδο ασφάλισης όπως υγείας, ζωής ή οικίας.

Κεφάλαιο 6

Συμπεράσματα

Ο σκοπός της παρούσας διπλωματικής εργασίας, ήταν να δείξει την πολύτιμη αξία της εξόρυξης πληροφοριών μέσα από μεγάλο σύνολο δεδομένων, κάνοντας χρήση σύγχρονων τεχνολογικών προγραμμάτων, βασισμένη στη θεωρία συστημάτων της Επιχειρηματικής Ευφυΐας. Για να κατανοηθεί λεπτομερώς η φύση αυτού του ζητήματος, εξερευνήθηκαν μέσω τεχνικών εξόρυξης πληροφορίας και οπτικοποιήσεων, του λογισμικού Tableau, οι συμπεριφορές και οι σχέσεις ασφαλιστικής εταιρείας με τους πελάτες της. Τα αποτελέσματα δίνουν πληροφορίες, αναλυτικά, για την συνολική και επιμέρους κατάσταση των ασφαλιστικών προϊόντων της εταιρείας, κατηγοριοποιούν τους πελάτες σε διάφορα καταναλωτικά γκρουπ, ανάλογα το σκοπό που αυτά εξετάζονται (ηλικία, περιοχή, επάγγελμα) και εμφανίζουν τάσεις-καταναλωτικές συνήθειες και συμπεριφορές. Με αυτές τις πληροφορίες, γίνεται πιο εύκολος ο εντοπισμός προβλημάτων και αδυναμιών, η ανάθεση στόχων και η δημιουργία στρατηγικών πλάνων που μπορεί να εφαρμόσει η διοίκηση των επιχειρήσεων.

Η χρήση του προγράμματος Tableau ανέδειξε τις ισχυρές δυνατότητες και τα πλεονεκτήματά του, τόσο σε πρώτο στάδιο με την ταχύτητα επεξεργασίας των δεδομένων της βάσης και την εξαγωγή συμπερασμάτων, όσο και σε δεύτερο στάδιο με την ευκολία κατανόησης των συμπερασμάτων, την απλότητα και την υψηλή αισθητική παρουσίαση

των οπτικοποιήσεων. Ακόμη, η μικρή παρέμβαση του ανθρώπινου παράγοντα στις διεργασίες και τους υπολογισμούς, έχει ως άμεσο αποτέλεσμα την ελαχιστοποίηση λάθους ή απροσεξίας.

Κλείνοντας, η παρούσα εργασία θα μπορούσε να αποτελέσει υλικό για μελλοντική έρευνα σε συνδυασμό με εφαρμογή της Μηχανικής Μάθησης, με βασικό σκοπό την επίτευξη αυτοματοποιημένων διεργασιών, υπολογισμών και οπτικοποιήσεων. Μελέτη και έρευνα μίας τέτοιας περίπτωσης, θα μπορούσε να εξασφαλίσει γρήγορες απαντήσεις που αφορούν τις καταναλωτικές συνήθειες των πελατών και ταυτόχρονα θα μείωνε αισθητά το φόρτο εργασίας του εργαζόμενου χρήστη.

Bibliography

- [1] AHMED, S. Applications of data mining in retail business. In *International Conference on Information Technology: Coding and Computing, 2004. Proceedings. ITCC 2004*. (2004), vol. 2, pp. 455–459 Vol.2.
- [2] AJIBADE, S. S., AND ADEDIRAN, A. An overview of big data visualization techniques in data mining. *International Journal of Computer Science and Information Technology Research* 4, 3 (2016), 105–113.
- [3] ANKERST, M., KEIM, D. A., AND KRIEGEL, H.-P. Circle segments: A technique for visually exploring large multidimensional data sets. In *Visualization* (1996).
- [4] BELLINGER, G., CASTRO, D., AND MILLS, A. Data, information, knowledge, and wisdom. 2004.
- [5] BORGIO, R., KEHRER, J., CHUNG, D. H. S., MAGUIRE, E., LARAMEE, R. S., HAUSER, H., WARD, M., AND CHEN, M. Glyph-based Visualization: Foundations, Design Guidelines, Techniques and Applications. In *Eurographics 2013 - State of the Art Reports* (2013), M. Sbert and L. Szirmay-Kalos, Eds., The Eurographics Association.
- [6] CHARTS, R. Scatter plot by group in r. <https://r-charts.com/correlation/scatter-plot-group/>.
- [7] DATAVIS. Animated transitions in tableau. <https://datavis.blog/2020/01/21/tableau-animated-transitions/>, 2020.
- [8] DAVID REINSEL, JOHN RYDNING, J. G. The digitization of the world. <https://www.seagate.com/files/www-content/our-story/trends/files/idc-seagate-dataage-whitepaper.pdf>, Nov. 2018.
- [9] DAVIES, T. Open data, democracy and public sector reform. *A look at open government data use from data. gov. uk* (2010), 1–47.
- [10] EUROPA, D. What is open data. <https://data.europa.eu/el/trening/what-open-data>, 2022.

- [11] EVOLYTICS. Tableau fundamentals: An introduction to dashboards and sharing. <https://evolytics.com/blog/tableau-fundamentals-introduction-dashboards-distribution/>.
- [12] FOUNDATION, O. K. Defining open in open data, open content and open knowledge. <https://opendefinition.org/>.
- [13] FOUNDATION, O. K. What is open. <https://okfn.org/opendata/>.
- [14] FRIEDMAN, A. Data and visual displays in the journal of ecology 1996–2016. *Information Visualization* 20, 1 (2021), 35–46.
- [15] GURIN, J. Big data and open data: How open will the future be? 2015.
- [16] HAJIRAHIMOVA, M., AND ISMAYILOVA, M. Big data visualization: Existing approaches and problems. *Problems of Information Technology* 9, 1 (2018), 65–74.
- [17] IBM. Operational risk management in the world of big data. <https://www.ibm.com/downloads/cas/716VKRP0>, 2014.
- [18] JAKLIČ, J. Assessing benefits of business intelligence systems – a case study. *Management: Journal of Contemporary Management Issues (mbuble@efst.hr); Vol.15 No.1* (01 2008).
- [19] JENIK, C. A minute on the internet. <https://www.statista.com/chart/25443/>, July 2021.
- [20] JETZEK, T., AVITAL, M., AND BJORN-ANDERSEN, N. Generating value from open government data. vol. 2. 2013.
- [21] KIMBLE CHRIS, M. G. Big data and business intelligence: Debunking the myths. *Global Business and Organizational Excellence* 35, 1 (2015), 23–34.
- [22] KIRKOS, E. Business intelligence and data mining [undergraduate textbook]. <http://hdl.handle.net/11419/1226>, 2015.
- [23] KIRKOS, E. Advanced topics in artificial intelligence. https://people.ieu.edu/~demos/Downloads/AI_ST_11_Business_I.pdf, 2021.
- [24] KRISHNAN, K. Introduction to big data. In *Data Warehousing in the Age of Big Data*. Morgan Kaufmann, 2013, pp. 3–14.
- [25] KRISHNAN, K. 1 - big data introduction. In *Building Big Data Applications*. Academic Press, 2020, pp. 1–16.
- [26] LAB, T. I. Show me how: Treemaps. <https://www.theinformationlab.co.uk/2015/02/10/show-treemaps/>, 2015.

- [27] LI NIU, JIE LU, Z. G. *Cognition-Driven Decision Support for Business Intelligence: Models, Techniques, Systems and Applications*. Springer Berlin Heidelberg, 2009, pp. 19–29.
- [28] MACQUEEN, J. Classification and analysis of multivariate observations. In *5th Berkeley Symp. Math. Statist. Probability* (1967), pp. 281–297.
- [29] MATTHAIOU, H. Big data. <https://nowmag.gr/big-data/>, Aug. 2020.
- [30] MICROSOFT. What is cloud computing. <https://azure.microsoft.com/en-us/resources/cloud-computing-dictionary/what-is-cloud-computing/#cloud-deployment-types>.
- [31] NEGASH, S. Business intelligence. *Communications of the Association for Information Systems* 13 (01 2004).
- [32] NUGENT, A., HALPER, F., HURWITZ, J., AND KAUFMAN, M. *Big Data For Dummies*. 2013.
- [33] PHYU, T. N. Survey of classification techniques in data mining. In *Proceedings of the international multiconference of engineers and computer scientists* (2009), vol. 1, Citeseer.
- [34] SAS. Data visualization techniques - from basics to big data with sas® visual analytics. https://www.sas.com/content/dam/SAS/en_us/doc/whitepaper1/data-visualization-techniques-106006.pdf, 2014.
- [35] SIEGEL, E. *Predictive analytics: The power to predict who will click, buy, lie, or die*. John Wiley & Sons, 2013.
- [36] STALIDIS G, K. D. *Data Management and Business Intelligence [Undergraduate textbook]*. Kallipos, Open Academic Editions., 2015.
- [37] TABLEAU. Understanding and using bar charts. <https://www.tableau.com/data-insights/reference-library/visual-analytics/charts/bar-charts>.
- [38] TABLEAU. Understanding and using line charts. <https://www.tableau.com/data-insights/reference-library/visual-analytics/charts/line-charts>.
- [39] TABLEAU. Understanding and using pie charts. <https://www.tableau.com/data-insights/reference-library/visual-analytics/charts/pie-charts>.
- [40] TABLEAU. What is tableau? <https://www.tableau.com/why-tableau/what-is-tableau>, 2022.
- [41] TABLEAU. The world’s leading analytics platform. <https://www.tableau.com>, 2022.

-
- [42] VERYKIOS V., KAGKLIS V., S. E. Data science through the r language [undergraduate textbook]. <http://hdl.handle.net/11419/2965>, 2015.
- [43] WIKIPEDIA. Byte. <https://el.wikipedia.org/wiki/Byte>, Nov. 2021.
- [44] WIKIPEDIA. Cloud computing. https://en.wikipedia.org/wiki/Cloud_computing, 2022.
- [45] WIKIPEDIA. Dashboard (business). [https://en.wikipedia.org/wiki/Dashboard_\(business\)](https://en.wikipedia.org/wiki/Dashboard_(business)), 2022.
- [46] WIKIPEDIA. Tableau software. https://en.wikipedia.org/wiki/Tableau_Software, 2022.