



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ  
ΤΜΗΜΑ ΟΡΓΑΝΩΣΗΣ ΚΑΙ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ  
ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ  
ΣΤΗ ΔΙΟΙΚΗΣΗ ΕΠΙΧΕΙΡΗΣΕΩΝ (ΜΒΑ)

**Διπλωματική Εργασία**

***Η καταναλωτική συμπεριφορά και οι σύγχρονοι τρόποι  
μελέτης της***

***Τα Μεγάλα Δεδομένα και η Ανάλυσή τους***

**Χαμπούλου Κυριακή του Γεωργίου**

**Πειραιάς, 2015**

*Αφιερώνεται στον σύζυγό μου  
που είναι κοντά μου και με στηρίζει σε κάθε μου βήμα*

# Η καταναλωτική συμπεριφορά και οι σύγχρονοι τρόποι μελέτης της Τα Μεγάλα Δεδομένα και η Ανάλυσή τους

Χαμοπούλου Κυριακή του Γεωργίου

Σημαντικοί όροι: Μεγάλα Δεδομένα, Αναλυτικά Μοντέλα, Επιχειρήσεις, Διορατικότητα, Αξία

## ΠΕΡΙΛΗΨΗ

Τα δεδομένα ήταν μία πραγματικότητα που υπήρχε πάντα, ωστόσο με την ανάπτυξη της τεχνολογίας (ηλεκτρονικοί υπολογιστές, έξυπνα τηλέφωνα (smartphones), βίντεο, φωτογραφίες, μέσα κοινωνικής δικτύωσης) αυξάνονται με ταχείς ρυθμούς. Οι ρυθμοί βέβαια αυτοί αναμένεται να αυξηθούν με την ανάπτυξη του Διαδικτύου των Πραγμάτων. Όλα αυτά τα δεδομένα, αν χρησιμοποιηθούν σωστά, με τη δημιουργία κατάλληλων αναλυτικών εργαλείων έχουν πάρα πολλά να προσφέρουν στις επιχειρήσεις, στην εκπαίδευση, στα ερευνητικά κέντρα και στην ανθρωπότητα γενικότερα. Λόγω του τεράστιου όγκου τους, της ταχύτητάς τους και της ποικιλίας τους, έχουν ονομαστεί Μεγάλα Δεδομένα.

Πρώτος στόχος της συγκεκριμένης εργασίας είναι η κατανόηση του τι είναι Μεγάλα Δεδομένα, ποια είναι τα χαρακτηριστικά τους, από πού παράγονται και ποιες ικανότητες χρειάζονται για να χρησιμοποιηθούν. Δεύτερος στόχος είναι η συλλογή των διαφόρων τεχνικών που έχουν αναπτυχθεί για την ανάλυσή τους. Τρίτος στόχος είναι οι τρόποι που μπορούν να χρησιμοποιηθούν ώστε να μετατραπούν τα δεδομένα σε διορατικότητα και γνώση, και κατ'επέκταση να προσφέρουν αξία σε μία επιχείρηση ή σε έναν οργανισμό. Όσον αφορά τη μεθοδολογία, η συγκεκριμένη εργασία αποτελεί μία βιβλιογραφική μελέτη που συνδυάζει άρθρα πάνω στο συγκεκριμένο θέμα από συμβουλευτικές εταιρείες (McKinsey, KPMG), από επιστημονικές εκδόσεις (Harvard Business Review), από μεγάλες εταιρείες που ασχολούνται με τα δεδομένα (IBM, SAP) καθώς και από εταιρείες έρευνας αγοράς όπως η IRI.

Τέλος, τα Μεγάλα Δεδομένα, η ανάλυσή τους και η ερμηνεία τους μπορούν να συμβάλλουν σημαντικά στην αύξηση της αξίας μίας επιχείρησης ή ενός οργανισμού. Τα στελέχη μπορούν να παίρνουν ακριβείς και γρήγορες αποφάσεις σχετικά με το ποια στρατηγική θα ακολουθήσουν, ώστε να πετύχουν τους στόχους τους. Ακόμη, μπορούν να ανακαλύπτουν τάσεις και μοτίβα συμπεριφοράς από τα δεδομένα που μέχρι πριν δεν ήταν ορατά και τα οποία μπορούν να τους οδηγήσουν σε καινοτομικές υπηρεσίες ή προϊόντα.

## Παράρτημα Β: Βεβαίωση Εκπόνησης Διπλωματικής Εργασίας



**ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ**  
**ΣΧΟΛΗ ΟΙΚΟΝΟΜΙΚΩΝ ΕΠΙΧΕΙΡΗΜΑΤΙΚΩΝ ΚΑΙ ΔΙΕΘΝΩΝ ΣΠΟΥΔΩΝ**  
**ΤΜΗΜΑ ΟΡΓΑΝΩΣΗΣ ΚΑΙ ΔΙΟΙΚΗΣΗΣ ΕΠΙΧΕΙΡΗΣΕΩΝ**  
**ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ**  
**ΣΤΗ ΔΙΟΙΚΗΣΗ ΕΠΙΧΕΙΡΗΣΕΩΝ ΓΙΑ ΣΤΕΛΕΧΗ**

### ΒΕΒΑΙΩΣΗ ΕΚΠΟΝΗΣΗΣ ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

(περιλαμβάνεται ως ξεχωριστή (δεύτερη) σελίδα στο σώμα της διπλωματικής εργασίας)

«Δηλώνω υπεύθυνα ότι η διπλωματική εργασία για τη λήψη του μεταπτυχιακού τίτλου σπουδών, του Πανεπιστημίου Πειραιώς, στη Διοίκηση Επιχειρήσεων για Στελέχη : E-MBA» με τίτλο: «Η καταναλωτική συμπεριφορά και οι σύγχρονοι τρόποι μελέτης της - Τα Μεγάλα Δεδομένα και η Ανάλυσή τους» έχει συγγραφεί από εμένα αποκλειστικά και στο σύνολό της. Δεν έχει υποβληθεί ούτε έχει εγκριθεί στο πλαίσιο κάποιου άλλου μεταπτυχιακού προγράμματος ή προπτυχιακού τίτλου σπουδών, στην Ελλάδα ή στο εξωτερικό, ούτε είναι εργασία ή τμήμα εργασίας ακαδημαϊκού ή επαγγελματικού χαρακτήρα.

Δηλώνω επίσης υπεύθυνα ότι οι πηγές στις οποίες ανέτρεξα για την εκπόνηση της συγκεκριμένης εργασίας, αναφέρονται στο σύνολό τους, κάνοντας πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο. Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του πτυχίου μου».

Υπογραφή Μεταπτυχιακού Φοιτητή/τριας.....

Όνοματεπώνυμο.....

Ημερομηνία.....

## Ευχαριστίες

*Το θέμα των Μεγάλων Δεδομένων και των Αναλυτικών Μοντέλων είναι ένα θέμα που βρίσκεται στο επίκεντρο τόσο στον τομέα της έρευνας αγοράς όσο και στους περισσότερους τομείς που χρησιμοποιούν δεδομένα για τη λήψη επιχειρηματικών αποφάσεων. Θα ήθελα να ευχαριστήσω την εταιρεία για την οποία εργάζομαι τα τελευταία 6,5 χρόνια, την IRI, η οποία με έχει βοηθήσει με τη συνεχή εκπαίδευση αλλά και τις νέες προκλήσεις να «διαβάζω» τα δεδομένα και να τα ερμηνεύω με τέτοιο τρόπο ώστε να προσθέτουν αξία στις συνεργαζόμενες επιχειρήσεις. Επιπλέον, θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου, κ.Μάρκο Τσόγκα που με βοήθησε να επιλέξω το συγκεκριμένο θέμα και να πάω ένα βήμα πιο πέρα, σε κάτι καινούργιο, που αποτελεί την τάση μίας νέας εποχής.*

# Περιεχόμενα

1.	Εισαγωγή .....	1
2.	Μεγάλα Δεδομένα .....	3
2.1	Χαρακτηριστικά .....	4
2.2	Μεγάλα Δεδομένα για καταναλωτές .....	7
2.3	Μεγάλα Δεδομένα για επιχειρήσεις.....	7
2.4	Έρευνα.....	10
2.5	Διαφορές των Μεγάλων Δεδομένων από τα Μικρά Δεδομένα .....	11
2.6	Τα πρόσωπα της επιστήμης των δεδομένων .....	13
2.7	Τύποι και ικανότητες της επιστήμης των δεδομένων .....	14
2.8	Επιστήμη δεδομένων χωρίς Μεγάλα Δεδομένα.....	17
2.9	Μεγάλα Δεδομένα χωρίς την επιστήμη των δεδομένων.....	20
2.10	Δεδομένα που παράγονται από τον άνθρωπο .....	22
2.11	Δεδομένα που παράγονται από μηχανές.....	23
2.12	Το Διαδίκτυο των Πραγμάτων - Internet of Things.....	23
2.13	Διαφορές μεταξύ των δεδομένων που παράγονται από ανθρώπους και των δεδομένων που παράγονται από μηχανές .....	24
2.14	Δομημένα δεδομένα .....	24
2.15	Αδόμητα δεδομένα .....	25
2.16	Ημι-δομημένα δεδομένα.....	25
2.17	Προκλήσεις σχετικά με την ποιότητα των δεδομένων .....	26
3	Αναλυτικά Μοντέλα για Μεγάλα Δεδομένα.....	28
3.1	Εισαγωγή .....	28
3.2	Στόχοι .....	30
3.3	Εργαλεία και τεχνικές ανάλυσης Μεγάλων Δεδομένων.....	30
3.4	Εμπόδια .....	38
3.5	Προκλήσεις .....	39
3.6	Νέες Τάσεις στην Ανάλυση των Μεγάλων Δεδομένων .....	42

4	Αναλυτικά μοντέλα .....	47
4.1	Brand Advantage - Marketing Mix Modelling .....	47
4.1.1	Γενικά για το Marketing Mix .....	48
4.1.2	Σχετικά με τα μέσα ενημέρωσης (Media).....	51
4.1.3	Planner .....	55
4.2	Ανάλυση καλαθιού .....	56
4.3	Βελτιστοποίηση κωδικολογίου.....	57
4.4	Demand Engineered Clustering .....	61
4.5	Συστάδες Καταστημάτων (Store Clustering).....	62
4.5.1	Ομαδοποίηση με βάση τις αγοραστικές περιστάσεις .....	65
4.6	Store Group Profilers .....	66
4.7	Μοντέλα προβλέψεων (Forecasting).....	67
4.8	Αγοραστικές περιστάσεις (Shopper Occasions) .....	70
4.8.1	Ομαδοποίηση με τη χρήση επιχειρηματικών κανόνων .....	70
4.8.2	Ομαδοποίηση με βάση τα πρότυπα κατανάλωσης.....	71
4.8.3	Ταξινόμηση νέων καλαθιών σε υπάρχουσες περιπτώσεις.....	71
4.9	Δίκτυα προϊόντων (Product Networks) .....	72
4.9.1	Τι είναι ένα δίκτυο προϊόντων .....	72
4.9.2	Δομή σε ένα δίκτυο προϊόντων.....	72
4.10	Κανόνες συσχέτισης (Association Rules) .....	74
5.	IRI - Analytics Center of Excellence (ACE).....	78
5.1	IRI & Analytics Center of Excellence (ACE).....	79
5.2	ACE - Analytics Center of Excellence.....	81
5.3	Μετάφραση των δεδομένων σε αξία.....	82
5.3.1	Μοντέλο SOLV (Situation, Opportunities, Limitations, Value) – Ανάλυση της κατάστασης, των ευκαιριών, των περιορισμών και της αξίας .....	83
5.3.2	Ανάλυση SWOT (Strengths, Weaknesses, Opportunities, Threats) .....	86
5.3.3	Προσέγγιση «Insights Factory» από τη McKinsey [4].....	90
5.3.4	Πέντε τρόποι για την μετατροπή των Μεγάλων Δεδομένων σε διορατικότητα και δράση [5].....	93

5.3.5 Από τα δεδομένα σε διορατικές λύσεις (insights): Το Blueprint για μία επιχείρηση [6] .....	94
5.3.6 Πηγαίνοντας πέρα από τα δεδομένα: Μετατροπή των δεδομένων σε διορατικές λύσεις (insights) και αξία [7] .....	97
6. Συμπεράσματα – Προτάσεις για περαιτέρω έρευνα .....	99
Βιβλιογραφία .....	101
Άλλες Πηγές .....	103



## 1. Εισαγωγή

Ο Bill Gates είπε κάποτε ότι “έχουμε υπερεκτιμήσει τις αλλαγές που θα γίνουν στα επόμενα δύο χρόνια και υποεκτιμήσει τις αλλαγές που θα γίνουν στα επόμενα δέκα”. Ο συνδυασμός της ψηφιακής τεχνολογίας, των προσδοκιών των καταναλωτών και οι κοινωνικο-οικονομικές αλλαγές έχουν αλλάξει τον τρόπο που οι καταναλωτές ψωνίζουν. Κατά συνέπεια οι προμηθευτές, οι βιομήχανοι και οι λιανέμποροι θα πρέπει να προσαρμοστούν αντίστοιχα. Για να το πετύχουν αυτό, θα πρέπει να συλλέξουν τα δεδομένα, να τα αναλύσουν και να καταλήξουν σε πλάνα δράσης που θα αυξήσουν την αξία της επιχείρησής τους.

Ο όγκος των δεδομένων, η ταχύτητα με την οποία παράγονται και η ποικιλία τους στην σημερινή εποχή, έχουν πάρει τόσο μεγάλες διαστάσεις που δημιουργήθηκε ένας νέος τομέας μελέτης τους, ο οποίος ονομάζεται Μεγάλα Δεδομένα. Τα Μεγάλα Δεδομένα και η ανάλυσή τους, τόσο στην Αμερική όσο και στην Ευρώπη, είναι ένας κλάδος που αναπτύσσεται συνεχώς και ταυτόχρονα χρησιμοποιείται από τις επιχειρήσεις για να βρουν κρυμμένες τάσεις πίσω από τα δεδομένα που θα τις βοηθήσουν να αναπτυχθούν περαιτέρω. Το θέμα που αναλύεται παρακάτω αφορά τα Μεγάλα Δεδομένα που προκύπτουν είτε από τους ανθρώπους είτε από τις μηχανές και την ανάλυσή τους.

Ο βασικός λόγος επιλογής του συγκεκριμένου θέματος είναι ότι αποτελεί μία από τις μεγαλύτερες τάσεις της εποχής στον τομέα της έρευνας αγοράς. Όλο και περισσότερες επιχειρήσεις χρησιμοποιούν την ανάλυση των δεδομένων στην λήψη αποφάσεων και σύμφωνα με την KPMG, είναι ο παίκτης που θα αλλάξει το παιχνίδι. Ο δεύτερος λόγος είναι, όπως περιγράφει και ο Thomas H. Davenport στο Harvard Business Review, ότι το επάγγελμα του επιστήμονα που θα αναλύει και θα ερμηνεύει τα Μεγάλα Δεδομένα θα είναι “η πιο ενδιαφέρουσα δουλειά του 21<sup>ου</sup> αιώνα” [26]. Αυτή τη στιγμή, οι άνθρωποι που ασχολούνται με την επιστήμη των δεδομένων είναι σε έλλειψη σε σχέση με τον συνεχώς αυξανόμενο ρυθμό χρήσης της. Σύμφωνα με το Παγκόσμιο Ινστιτούτο της McKinsey, μόνο οι Ηνωμένες Πολιτείες της Αμερικής αντιμετωπίζουν ένα έλλειμμα 140.000 με 190.000 ανθρώπους με αναλυτικές ικανότητες.

Η σημασία των Μεγάλων Δεδομένων για τον επιχειρηματικό κόσμο είναι τεράστια, καθώς τα στελέχη των επιχειρήσεων μπορούν να λαμβάνουν αποφάσεις βασιζόμενοι σε συγκεκριμένα δεδομένα και σε πλάνα δράσης που θα έχουν προκύψει από την ανάλυσή τους. Με τα μοντέλα πρόβλεψης, τα οποία δίνουν τη δυνατότητα δημιουργίας σεναρίων, θα ξέρουν τα αποτελέσματα από το κάθε πλάνο δράσης και θα μπορούν να επιλέγουν με περισσότερη ασφάλεια ποια στρατηγική θα πρέπει να ακολουθήσουν για να πετύχουν συγκεκριμένους στόχους. Επιπλέον, μέσα από την ανάλυση και ερμηνεία των Μεγάλων Δεδομένων θα έχουν τη δυνατότητα να ανακαλύψουν νέες ευκαιρίες και ανταγωνιστικά πλεονεκτήματα. Σύμφωνα με μία έρευνα της KPMG σε στελέχη επιχειρήσεων, το 86% αυτών απάντησαν ότι τα αναλυτικά μοντέλα τους βοήθησαν στην πιο γρήγορη λήψη αποφάσεων, το 67% ότι συνέβαλλαν να μειώσουν τον επιχειρηματικό κίνδυνο και το 80% είπε ότι παίρνουν πιο ακριβείς αποφάσεις [22].

Όσον αφορά τη μεθοδολογία είναι μία βιβλιογραφική μελέτη που συνδυάζει περιγραφή των μεγάλων δεδομένων, των αναλυτικών μοντέλων για μεγάλα δεδομένα και των μελλοντικών τάσεων. Παράλληλα, περιγράφονται τα αναλυτικά μοντέλα που προσφέρει η IRI μέσω της επίλεκτης ομάδας επιστημόνων που διαθέτει στην Αθήνα, την ACE (Athens Center of Excellence), η οποία εξυπηρετεί επιχειρήσεις κυρίως στην Ευρώπη και στην Αμερική, αλλά και στον υπόλοιπο κόσμο. Τέλος, περιγράφονται ορισμένα μοντέλα που χρησιμοποιούνται για την μετατροπή των δεδομένων σε διορατικότητα και αξία.

## 2. Μεγάλα Δεδομένα

Τα Μεγάλα Δεδομένα αναφέρονται σε δεδομένα τα οποία λόγω του μεγέθους τους, της ταχύτητας ή της μορφής τους δεν είναι εύκολο να αποθηκευτούν, να διαχειριστούν ή να αναλυθούν με τις παραδοσιακές μεθόδους, όπως τα λογιστικά φύλλα (Excel), τις σχεσιακές βάσεις δεδομένων ή τα κοινά λογισμικά στατιστικής. Τα Μεγάλα Δεδομένα είναι ένας διφορούμενος και σχετικός όρος και ίσως είναι καλύτερα να οριστεί αντίστροφα, δηλαδή να περιγράψουμε τι δεν είναι Μεγάλα Δεδομένα. Δεν είναι τα κανονικά στοιχεία, δεν είναι τα συνηθισμένα δεδομένα και δεν είναι κάτι με το οποίο ένας έμπειρος αναλυτής μπορεί να είναι έτοιμος να διαχειριστεί. Με άλλα λόγια, Μεγάλα Δεδομένα είναι δεδομένα που δεν ταιριάζουν καλά σε ένα τυπικό αναλυτικό παράδειγμα. Δεν θα μπορούσαν να ενσωματωθούν σωστά στις σειρές και στις στήλες ενός υπολογιστικού φύλλου. Δεν μπορούν να αναλυθούν με συμβατικά μοντέλα πολλαπλής παλινδρόμησης και κατά πάσα πιθανότητα δεν μπορούν να αποθηκευτούν σε έναν κανονικό σκληρό δίσκο υπολογιστή οικιακής χρήσης.

Παράλληλα, ένας άλλος τρόπος περιγραφής των Μεγάλων Δεδομένων είναι χρησιμοποιώντας τα τρία V, τα οποία είναι ο Όγκος (Volume), η Ταχύτητα (Velocity), και η ποικιλία (Variety). Σε μια έρευνα του 2001 [1] και σε σχετικές διαλέξεις, ο αναλυτής Doug Laney του META Group (τώρα Gartner) ορίζει τις όλο και αυξανόμενες προκλήσεις και ευκαιρίες της ανάπτυξης των δεδομένων με τρεις διαστάσεις, με την αύξηση του όγκου (ποσότητα δεδομένων), με την ταχύτητα (ταχύτητα των δεδομένων μέσα και έξω από έναν οργανισμό), και την ποικιλία (εύρος των τύπων δεδομένων και των πηγών). Η Gartner, και τώρα ένα μεγάλο μέρος της βιομηχανίας, συνεχίζουν να χρησιμοποιούν το μοντέλο των "3Vs" για την περιγραφή των μαζικών δεδομένων [2]. Το 2012, η Gartner αναθεώρησε τον ορισμό ως εξής: "Μεγάλα Δεδομένα είναι οι μεγάλου όγκου, υψηλής ταχύτητας, και / ή μεγάλης ποικιλίας πληροφορίες-περιοριστικά στοιχεία που απαιτούν νέες μορφές επεξεργασίας για να καταστεί δυνατή η ενισχυμένη λήψη αποφάσεων, η ανακάλυψη της διορατικότητας (insights) και η βελτιστοποίηση των διαδικασιών. Επιπλέον, ένα νέο V που προστίθεται από ορισμένες οργανώσεις, είναι η ακρίβεια (Veracity)"[3].

Αν και ο ορισμός της Gartner (3Vs) εξακολουθεί να χρησιμοποιείται ευρέως, η αυξανόμενη ωρίμανση της έννοιας ενισχύει μια πιο υγιή διαφορά μεταξύ των Μεγάλων Δεδομένων και της Επιχειρηματικής Ευφυΐας (Business Intelligence), όσον αφορά τα

δεδομένα και τη χρήση τους: [4]. Η Επιχειρηματική Ευφυΐα (Business Intelligence) χρησιμοποιεί περιγραφική στατιστική και στοιχεία με υψηλή πυκνότητα πληροφοριών για να μετρήσει πράγματα, να ανιχνεύσει τις τάσεις. Από την άλλη πλευρά, τα Μεγάλα Δεδομένα χρησιμοποιούν επαγωγική στατιστική και έννοιες μη γραμμικών συστημάτων [5] για να αναφερθούν σε αναλύσεις (παλινδρομήσεις, μη γραμμικές σχέσεις, και σχέσεις αιτιώδους συνάφειας) από μεγάλα σύνολα δεδομένων με χαμηλή πυκνότητα πληροφοριών [6] για να αποκαλύψουν σχέσεις, εξαρτήσεις και να κάνουν προβλέψεις αποτελεσμάτων και συμπεριφορών. [5] [7]

Ένας πιο πρόσφατος, συναινετικός ορισμός αναφέρει ότι «Τα Μεγάλα Δεδομένα αντιπροσωπεύουν τα περιουσιακά στοιχεία - πληροφορίες που χαρακτηρίζονται από τόσο μεγάλο όγκο, υψηλή ταχύτητα και μεγάλη ποικιλία που για να μετατραπούν σε Αξία απαιτούν ειδική τεχνολογία και αναλυτικές μεθόδους». [8]

## 2.1 Χαρακτηριστικά [28]

Τα Μεγάλα Δεδομένα μπορούν να περιγραφούν από τα παρακάτω χαρακτηριστικά:

- Όγκος (Volume)

Με τον απλούστερο δυνατό ορισμό «Μεγάλα Δεδομένα» είναι τα δεδομένα που είναι πάρα πολύ μεγάλα σε όγκο ώστε να επεξεργαστούν από έναν απλό υπολογιστή. Προφανώς αυτό είναι πολύ σχετικό, γιατί αυτό που είναι μεγάλο για ένα σύστημα σε μία συγκεκριμένη χρονική στιγμή, μπορεί να είναι κάτι συνηθισμένο για ένα άλλο σύστημα σε μία άλλη χρονική στιγμή. Ως γενική παρατήρηση του νόμου του Moore, είναι ότι στην Επιστήμη των Υπολογιστών η χωρητικότητα και η απόδοση των υπολογιστών διπλασιάζεται περίπου κάθε δύο χρόνια. Παραδείγματα που το αποδεικνύουν είναι η αλλαγή του μεγέθους των υπολογιστικών φύλλων του Excel, οι φωτογραφίες και τα βίντεο. Μερικοί άνθρωποι το αποκαλούν πολλά δεδομένα, και εννοούν το ίδιο πράγμα, ότι έχουμε πολλά στοιχεία. Απλά τώρα έχουμε πολύ περισσότερα από αυτό που πιστεύουν, και σε αυτό έχει παίξει ρόλο η ταχύτητα και η ποικιλία.

Η ποσότητα των παραγόμενων δεδομένων είναι σημαντική σε αυτό το πλαίσιο. Το μέγεθος των δεδομένων καθορίζει την αξία και τις δυνατότητες των εξεταζόμενων δεδομένων, καθώς και αν μπορούν πράγματι να θεωρηθούν Μεγάλα Δεδομένα ή όχι.

Οι λέξεις «Μεγάλα Δεδομένα» από μόνες τους περιέχουν έναν όρο που σχετίζεται με το μέγεθος, και ως εκ τούτου με το συγκεκριμένο χαρακτηριστικό.

- Ποικιλία (Variety)

Η ποικιλία αποτελεί ένα σημαντικό χαρακτηριστικό των Μεγάλων Δεδομένων και είναι κάτι που οι αναλυτές θα πρέπει να γνωρίζουν. Η ποικιλία βοηθά τους ανθρώπους που αναλύουν τα δεδομένα ώστε να τα συνδέσουν και να τα χρησιμοποιήσουν αποτελεσματικά προς όφελός τους.

Η ποικιλία δεν περιγράφει μόνο τις γραμμές και τις στήλες ενός καλά καθορισμένου συνόλου δεδομένων, σε ένα υπολογιστικό φύλλο για παράδειγμα. Αντίθετα, περιλαμβάνει πολλά διαφορετικά φύλλα σε πολλές διαφορετικές μορφές. Μπορεί να είναι ένα αδόμητο κείμενο, όπως βιβλία, απόψεις σε ιστότοπους (blog), σχόλια σε άρθρα ειδήσεων, ή tweets από το Twitter. Οι ερευνητές εκτιμούν ότι το 80% των στοιχείων των επιχειρήσεων μπορεί να έχουν μη δομημένη μορφή. Επιπλέον, οποιοδήποτε είδος δεδομένων που δεν ταιριάζει σε μία παραδοσιακή, δομημένη, σχεσιακή βάση δεδομένων μπορεί να δημιουργήσει πολλές σημαντικές αναλυτικές προκλήσεις. Τέλος, η ποικιλία είναι ο μεγαλύτερος παράγοντας που οδηγεί τις εταιρείες σε λύσεις Μεγάλων Δεδομένων.

- Ταχύτητα (Velocity)

Ταχύτητα στο πλαίσιο αυτό σημαίνει πόσο γρήγορα παράγονται και υφίστανται επεξεργασία τα δεδομένα για να ικανοποιήσουν τις απαιτήσεις και τις προκλήσεις που βρίσκονται στο δρόμο της ανάπτυξης.

Σε μια συμβατική επιστημονική έρευνα μπορεί να πάρει μήνες για να γίνει η συλλογή των δεδομένων για 100 περιπτώσεις, εβδομάδες για να γίνει η ανάλυση και χρόνια για να δημοσιευτεί η έρευνα. Παράλληλα με τον χρόνο που χρειάζεται για τα παραπάνω στάδια, είναι και το γεγονός ότι είναι μία μελέτη στατική, δεν είναι δυναμική. Για παράδειγμα, ίσως το πιο γνωστό σύνολο δεδομένων για τη διδασκαλία της στατιστικής ανάλυσης όσον αφορά την ομαδοποίηση (clustering analysis) [9], είναι τα δεδομένα που έχουν συγκεντρωθεί από τον Edgar Andersen και αναλύονται από τον Robert Fisher. Το συγκεκριμένο σύνολο δεδομένων εξακολουθεί να χρησιμοποιείται κάθε μέρα και δεν έχει αλλάξει τα τελευταία ογδόντα χρόνια. Στην άλλη άκρη της κλίμακας, η κοινωνική πλατφόρμα μέσων μαζικής ενημέρωσης "Twitter" παράγει μια τεράστια

ποσότητα νέων δεδομένων που είναι περίπου 6.000 tweets ανά δευτερόλεπτο ή 500 εκατομμύρια tweets ανά ημέρα και περίπου 200 δισεκατομμύρια tweets ανά έτος. Αυτό το είδος της συνεχούς εισόδου των δεδομένων που ονομάζεται επίσης Ροή Δεδομένων Αισθητήρων (Streaming Sensor Data) παρουσιάζει ιδιαίτερες προκλήσεις για τους αναλυτές, δεδομένου ότι τα στοιχεία αυτά είναι «ρευστά».

- Μεταβλητότητα (Variability)

Αυτή η διάσταση αναφέρεται στην ασυνέπεια που μπορεί να έχουν τα δεδομένα κατά καιρούς, η οποία παρακωλύει τη διαδικασία διαχείρισης των δεδομένων με αποτελεσματικό τρόπο.

- Ακρίβεια (Veracity)

Η ποιότητα των δεδομένων μπορεί να ποικίλλει σε μεγάλο βαθμό. Η ακριβής ανάλυσή τους εξαρτάται από την εμπιστοσύνη που μπορεί να έχει κάποιος στην πηγή των πληροφοριών.

- Πολυπλοκότητα

Η διαχείριση των δεδομένων μπορεί να είναι πολύ περίπλοκη, ειδικά όταν μεγάλος όγκος δεδομένων προέρχεται από διαφορετικές πηγές. Τα δεδομένα πρέπει να συνδέονται και να συσχετίζονται μεταξύ τους, έτσι ώστε οι χρήστες να μπορούν να κατανοήσουν τις πληροφορίες που υποτίθεται ότι μεταφέρουν.

Το τελευταίο ζήτημα είναι, αν θα πρέπει να υφίστανται και τα τρία Vs (Volume, Velocity και Variety) ταυτόχρονα ή ένα από τα τρία αρκεί για να περιγραφεί ένα σύνολο δεδομένων ως Μεγάλα Δεδομένα. Είναι λογικό ότι αν συνυπάρχουν και τα τρία τότε έχουμε Μεγάλα Δεδομένα, αλλά κάθε ένα από τα τρία, από μόνο του μπορεί να είναι σε τέτοιο βαθμό που να μην μπορεί να χρησιμοποιηθεί μία τυποποιημένη προσέγγιση. Κατά συνέπεια, τα Μεγάλα Δεδομένα μπορεί να παρουσιάσουν μια σειρά από ιδιαίτερες προκλήσεις.

## 2.2 Μεγάλα Δεδομένα για καταναλωτές

Τις περισσότερες φορές που οι άνθρωποι μιλούν για Μεγάλα Δεδομένα, μιλούν για αυτά από την «εμπορική πλευρά τους». Σχετικά με το πώς οι επιχειρήσεις μπορούν να τα χρησιμοποιήσουν για στρατηγικές διαφήμισης ή μάρκετινγκ. Ωστόσο, ένα πολύ σημαντικό μέρος των Μεγάλων Δεδομένων χρησιμοποιείται από τους καταναλωτές. Το ενδιαφέρον γεγονός είναι ότι οι αλγόριθμοι που χρησιμοποιούνται και όλη η εξεζητημένη επεξεργασία είναι σχεδόν διαφανής για τους καταναλωτές. Τα αποτελέσματα είναι τόσο καθαρά, ώστε οι καταναλωτές να παίρνουν μόνο ένα μικρό μέρος πληροφοριών, αλλά αυτό το μέρος να είναι αυτό ακριβώς που χρειάζονται.

Οι μεγάλοι οργανισμοί είχαν αξιοποιήσει την δύναμη της ανάλυσης των δεδομένων για κάποιο χρονικό διάστημα. Όμως, οι εταιρείες παροχής υπηρεσιών προς τον καταναλωτή βρίσκουν νέους τρόπους για να χρησιμοποιήσουν την επιχειρηματική ευφυΐα (Business Intelligence), ώστε να ωφεληθούν τα ίδια τα άτομα. Ιστοσελίδες όπως Amazon, Pandora και Netflix ξεκίνησαν αυτή την τάση, χρησιμοποιώντας εξελιγμένη επιχειρηματική ευφυΐα (Business Intelligence) για να καταλάβουν και να προτείνουν πράγματα που ενδεχομένως θα ήθελαν να αγοράσουν οι καταναλωτές, να ακούσουν ή να παρακολουθήσουν.

Ένα πράγμα που κάνει όλα αυτά να είναι εφικτά, είναι η αυξανόμενη διαθεσιμότητα των μεγάλων δημόσιων και ιδιωτικών πηγών πληροφόρησης. Οι κρατικές υπηρεσίες και εταιρείες όπως το Facebook Inc., η Google Inc. και το Twitter Inc. προσφέρουν περιβάλλον προγραμματισμού εφαρμογών API (application programming interface) που επιτρέπει σε άλλους κατασκευαστές λογισμικού να έχουν πρόσβαση και να χρησιμοποιούν τα δεδομένα τους. Παρόλο που οι καταναλωτές ανησυχούν για την επίδραση που μπορεί να έχει στην ιδιωτική τους ζωή η δημοσίευση των προσωπικών τους πληροφοριών, πολλοί βρίσκουν τρόπους για να επωφεληθούν από τις νέες, εύκολα προσβάσιμες, υπηρεσίες δεδομένων.

## 2.3 Μεγάλα Δεδομένα για επιχειρήσεις

Τα Μεγάλα Δεδομένα μπορούν μεν να δώσουν σημαντικές ανέσεις και λειτουργικότητα στους καταναλωτές, αλλά από την άλλη πλευρά στον κόσμο των επιχειρήσεων

ανατρέπουν εκ των βάθρων τον τρόπο με τον οποίο οι άνθρωποι εμπορεύονται. Οι περισσότεροι άνθρωποι έχουν εντοπίσει τα Μεγάλα Δεδομένα όσον αφορά το εμπόριο στα αποτελέσματα της Google, στις πρόσθετες αναζητήσεις. Κάθε φορά που γίνεται μια αναζήτηση στο Google ή σε μια παρόμοια μηχανή αναζήτησης, εμφανίζεται μεν το αποτέλεσμα, αλλά εμφανίζονται ταυτόχρονα και διαφημίσεις. Οι διαφημίσεις εμφανίζονται με βάση τα κριτήρια και τις πληροφορίες της Google για τον τελικό χρήστη. Η Google τοποθετεί διαφημίσεις με βάση το προφίλ του χρήστη και εμφανίζει τα στοιχεία στα οποία ο χρήστης πιθανότατα θα ανταποκριθεί. Αυτό επιτυγχάνεται με την επεξεργασία μίας πολύ μεγάλης ποσότητας των διαθέσιμων δεδομένων.

- Marketing προβλέψεων (Predictive Marketing)

Μια άλλη ενδιαφέρουσα περιοχή είναι το Μάρκετινγκ προβλέψεων, όπου τα Μεγάλα Δεδομένα χρησιμοποιούνται για να βοηθήσουν τον χρήστη να αποφασίσει που θα πάει προτού πραγματικά να πάει εκεί. Χρησιμοποιείται για να προβλέψει έναν αριθμό πραγμάτων που συμπεριλαμβάνουν:

ο Τα σημαντικά γεγονότα της ζωής - Όπως την αποφοίτηση, τον γάμο, μία νέα θέση εργασίας, τη γέννηση ενός παιδιού, ή οποιοδήποτε συμβάν έχει να κάνει με μια σειρά εμπορικών συναλλαγών.

ο Εξετάζει τη συμπεριφορά του καταναλωτή - Οι εταιρείες μπορούν να παρακολουθήσουν πόσο συχνά ένας χρήστης συνδέεται σε μια ιστοσελίδα, ποιες πιστωτικές κάρτες χρησιμοποιεί, ή πόσο συχνά βλέπει συγκεκριμένα στοιχεία πριν τη μετάβασή του σε κάτι άλλο. Έτσι, μία τεράστια ποσότητα πληροφοριών που είναι ήδη διαθέσιμες μπορούν να επεξεργαστούν από τις εταιρείες.

ο Χρησιμοποιεί δημογραφικές πληροφορίες - Αυτό το είδος των πληροφοριών μπορεί να περιλαμβάνει την ηλικία, την οικογενειακή κατάσταση, τον αριθμό των παιδιών, την διεύθυνση κατοικίας, την απόσταση της διεύθυνσης του σπιτιού από ένα συγκεκριμένο κατάστημα, τον εκτιμώμενο μισθό, τις πιστωτικές κάρτες που χρησιμοποιεί ο χρήστης, και τι είδους ιστοσελίδες ο χρήστης επισκέπτεται πιο συχνά. Όλη αυτή η πληροφόρηση είναι διαθέσιμη με τον ένα ή με τον άλλο τρόπο σε μία εταιρεία που προσπαθεί να δώσει σε αυτές τις πληροφορίες μορφή, να τις κάνει προϊόν ή υπηρεσία.

ο Μπορεί να αγοράσει περισσότερα δεδομένα - Οι εταιρείες μπορούν επίσης να αγοράσουν πρόσθετα στοιχεία για επεξεργασία, συμπεριλαμβανομένης της



εθνικότητας, προηγούμενες εργασίες, αν ένα άτομο έχει χρεοκοπήσει ή είναι διαζευγμένο, εκπαίδευση, κ.λπ. Υπάρχει ένα τεράστιο ποσό πληροφοριών που είναι δυνητικά διαθέσιμο.

Με τη χρήση αυτού του είδους των πληροφοριών είναι δυνατό μια εταιρεία να προβλέψει ότι ένα άτομο είναι έτοιμο να αγοράσει ένα νέο σπίτι. Με την αγορά ενός νέου σπιτιού θα αγοράσει και πολλά άλλα συναφή προϊόντα που συνδέονται με αυτό. Επομένως οι εταιρείες έχουν τη δυνατότητα να συνδεθούν, και να παρουσιάσουν τις λύσεις τους στον χρήστη.

#### • Πρόβλεψη Τάσεων

Ένας άλλος σημαντικός τομέας είναι όταν οι εταιρείες θέλουν να προβλέψουν τις τάσεις. Μια ενδιαφέρουσα περιοχή που αυτό μπορεί να εφαρμοστεί είναι η μόδα. Εταιρείες όπως η EDITD αναλύουν δεδομένα προκειμένου να προβλέψουν τις νέες τάσεις, σε σημείο που να μπορούν να πουν ποιοι θα είναι λιανέμποροι, ποια θα είναι τα πιο δημοφιλή χρώματα, τα στυλ, και τα εμπορικά σήματα, καθώς και πότε πρόκειται να είναι αυτά δημοφιλή. Μπορούν επίσης να διαμορφώσουν την τιμολογιακή τους πολιτική με βάση την ανάλυσή τους. Αυτό το είδος πληροφοριών έχει τεράστια αξία για τις εταιρείες που πωλούν προϊόντα, και η αξία αυτή μπορεί να επιτευχθεί με τη χρήση των Μεγάλων Δεδομένων.

#### • Ανίχνευση Απάτης

Ένας άλλος σημαντικός τομέας εφαρμογής των Μεγάλων Δεδομένων είναι η ανίχνευση απάτης. Η απάτη είναι μια τεράστια βιομηχανία, όπου οι λιανέμποροι χάνουν μεγάλα χρηματικά ποσά κάθε χρόνο λόγω της ηλεκτρονικής απάτης. Επίσης υπάρχει και η ασφαλιστική απάτη, όπου χάνονται ακόμη μεγαλύτερα ποσά. Κατά συνέπεια, η απάτη είναι ένα μεγάλο θέμα και υπάρχουν ορισμένα πράγματα που οι εταιρείες μπορούν να κάνουν για να μειώσουν την απάτη σε διαδικτυακές συναλλαγές. Αυτά είναι:

ο Το σημείο πώλησης - Αυτό σημαίνει συγκεκριμένα το πώς γίνεται η αγορά, αν είναι σε απευθείας σύνδεση, και ποια ιστοσελίδα χρησιμοποιείται.

ο Γεωγραφική τοποθεσία και IP διεύθυνση - Εδώ οι εταιρείες μπορούν να ελέγξουν που βρίσκεται ο πελάτης ως φυσική υπόσταση οπουδήποτε στον κόσμο ή μπορεί να

ελεγχθεί η διεύθυνση IP και να ανακτηθούν πληροφορίες σχετικά με τον υπολογιστή του χρησιμοποιήθηκε για την πρόσβαση στην ιστοσελίδα.

ο Χρόνος Σύνδεσης - Ο έλεγχος μπορεί επίσης να γίνει με το χρόνο σύνδεσης. Για παράδειγμα, εάν μια αγορά γίνεται στις τρεις το πρωί και ο χρήστης μπαίνει την ιστοσελίδα πριν από τις έντεκα, τότε αυτό θα μπορούσε να είναι καχύποπτο και να θεωρηθεί ότι έγινε μια απάτη.

ο Βιομετρικά - Οι επιχειρήσεις μπορούν επίσης να ελέγξουν τη χρήση βιομετρικών στοιχείων, δεδομένου ότι ο τρόπος που ένας χρήστης κινεί το ποντίκι ή κρατά ένα κινητό τηλέφωνο είναι μοναδικά και μπορούν να διακριθούν (εδώ χρησιμοποιούνται επιταχυνσιόμετρα, μοτίβα, μπορεί να υπολογιστεί η απόσταση του τηλεφώνου από το έδαφος). Έτσι, με τη χρήση των Μεγάλων Δεδομένων μία δυνητική απάτη μπορεί να αποφευχθεί.

## 2.4 Έρευνα

Τα Μεγάλα Δεδομένα, όπως αναφέρθηκε και παραπάνω, μπορούν να παίξουν μεγάλο ρόλο στη ζωή μεμονωμένων ανθρώπων, όπως και στις επιχειρήσεις. Από την άλλη πλευρά όμως, δεν μπορεί να αγνοηθεί το γεγονός ότι έχουν φέρει την επανάσταση σε πολλές πτυχές της επιστήμης και της έρευνας. Υπάρχουν πολλά παραδείγματα για το πόσο τα Μεγάλα Δεδομένα έχουν επηρεάσει την επιστημονική έρευνα, συμπεριλαμβανομένων των παρακάτω:

ο Τάσεις της γρίπης μέσω της Google - Οι επιστήμονες είναι σε θέση να προβλέψουν μέσα από μοτίβα αναζητήσεων, όπου χρησιμοποιούνται σχετικές λέξεις με τη γρίπη αν θα υπάρξουν και πότε ξεσπάσματα γρίπης στις Ηνωμένες Πολιτείες πολύ πριν συμβούν.

ο Plos - Υπολογιστική Βιολογία (Computational Biology) - Ένα άλλο παράδειγμα για τις τάσεις της γρίπης, όπου εντοπίζονται με μεγαλύτερη ακρίβεια, είναι με τη χρήση των αναζητήσεων στη Wikipedia.

ο Εθνικό Ινστιτούτο Υγείας (Brain Initiative) - Το Εθνικό Ινστιτούτο Υγείας δημιούργησε το (Brain Initiative) σαν έναν τρόπο να αντλήσει έναν τεράστιο αριθμό απεικονίσεων του εγκεφάλου και να δημιουργήσει μία πλήρη χαρτογράφηση της λειτουργίας του.

ο Κέπλερ (Kepler) – Επιπρόσθετα, το διαστημικό τηλεσκόπιο Κέπλερ (Kepler) της NASA είναι σε μια αποστολή να βρει εξω-πλανήτες ή πλανήτες εκτός του ηλιακού μας συστήματος. Μέχρι τώρα έχει βρει πάνω από χίλιους πλανήτες, και πάνω από τέσσερις χιλιάδες υποψηφίους.

ο American Psychological Association – Η ψυχολογική έρευνα έχει επίσης επηρεαστεί από τα Μεγάλα Δεδομένα. Μια μελέτη που έγινε σχετικά με τις προσωπικότητες στις Ηνωμένες Πολιτείες ήταν σε θέση να εντοπίσει ομάδες προσωπικοτήτων σε διαφορετικές περιοχές.

ο Βιβλία Google – Η Google έχει σαρώσει βιβλία που εκδόθηκαν τα τελευταία εκατό χρόνια προκειμένου να γίνουν ψηφιακά. Πάνω σε αυτά έγινε μια έρευνα για να βρεθεί η αλλαγή του τρόπου χρήσης των λέξεων με την πάροδο του χρόνου.

Ολοκληρώνοντας αυτή την ενότητα, τα Μεγάλα Δεδομένα με τον όγκο των πληροφοριών που είναι διαθέσιμος, την ποικιλία των πληροφοριών που μπορούν συνδυαστούν και την ταχύτητα ειδικά στις περιπτώσεις όπως η γρίπη, όπου υπάρχει συνεχής ροή δεδομένων, γίνεται καλή χρήση τους για την έρευνα και την επιστημονική πρόοδο.

## 2.5 Διαφορές των Μεγάλων Δεδομένων από τα Μικρά Δεδομένα

Όπως αναφέρθηκε ήδη τα Μεγάλα Δεδομένα χαρακτηρίζονται από περισσότερα από τρία Vs, Όγκο (Volume), Ταχύτητα (Veracity), Ποικιλία (Variety). Ωστόσο, υπάρχουν αρκετές πρακτικές διαφορές των μεγάλων στοιχείων από τα μικρά.

Όπως ανέφερε ο Jewels Berman [10] υπάρχουν δέκα κύριες διαφορές μεταξύ των Μεγάλων και Μικρών δεδομένων:

ο **Στόχοι** – Τα Μικρά Δεδομένα συλλέγονται συνήθως για έναν συγκεκριμένο στόχο. Τα Μεγάλα Δεδομένα από την άλλη πλευρά μπορεί αρχικά να συλλέγονται για έναν συγκεκριμένο στόχο και στη συνέχεια να πάρουν απροσδόκητες κατευθύνσεις.

ο **Τοποθεσία** - Τα Μικρά Δεδομένα είναι συνήθως σε ένα μέρος και συχνά σε ένα ενιαίο ηλεκτρονικό αρχείο. Τα Μεγάλα Δεδομένα από την άλλη πλευρά μπορεί να υπάρχουν σε πολλαπλά αρχεία, διακομιστές (servers) ή σε υπολογιστές σε διαφορετικές γεωγραφικές τοποθεσίες.

ο **Δομή δεδομένων και περιεχομένου** - Τα Μικρά Δεδομένα είναι συνήθως δομημένα (σε υπολογιστικό φύλλο Excel, σε γραμμές και στήλες). Τα Μεγάλα Δεδομένα από την άλλη πλευρά μπορεί να είναι αδόμητα, μπορεί να έχουν πολλά αρχεία και μορφές, και μπορούν να συνδεθούν με διάφορους τρόπους.

ο **Προετοιμασία Δεδομένων** - Τα Μικρά Δεδομένα συνήθως προετοιμάζονται από τον τελικό χρήστη για διαφορετικούς σκοπούς. Τα Μεγάλα Δεδομένα, συχνά προετοιμάζονται από μια ομάδα ανθρώπων, θα αναλυθούν από μία δεύτερη ομάδα ανθρώπων και θα χρησιμοποιηθούν από μία τρίτη ομάδα ανθρώπων. Μετά από αυτό, μπορεί να υπάρχουν για διαφορετικούς σκοπούς.

ο **Διάρκεια ζωής** - Τα Μικρά Δεδομένα συνήθως διατηρούνται για ένα συγκεκριμένο χρονικό διάστημα μετά το τέλος ενός έργου. Τα Μεγάλα Δεδομένα, επειδή κάθε έργο έχει μεγάλο κόστος, μπορούν να χρησιμοποιηθούν και για άλλα έργα είτε ως νέα δεδομένα, είτε να προστεθούν επιπλέον μεταβλητές και αρχεία. Έτσι, ένα έργο με Μεγάλα Δεδομένα μπορεί να έχει σημαντικά μεγαλύτερη διάρκεια ζωής από ό, τι ένα έργο μικρών δεδομένων.

ο **Μετρήσεις** - Τα Μικρά Δεδομένα συνήθως μετριοούνται με ένα ενιαίο πρωτόκολλο ή ένα σύνολο μονάδων, και αυτό γίνεται συνήθως την ίδια στιγμή. Με τα Μεγάλα Δεδομένα από την άλλη πλευρά, μπορεί να υπάρχουν άνθρωποι σε διαφορετικά μέρη ή διαφορετικούς χρόνους ή οργανισμούς ή διαφορετικές χώρες και ακόμα και διαφορετικά πρωτόκολλα. Έτσι, προκειμένου να μετρηθούν τα Μεγάλα Δεδομένα θα πρέπει να χρησιμοποιηθούν διαφορετικά πρωτόκολλα.

ο **Επαναληψιμότητα** - Μικρά σύνολα δεδομένων μπορούν συνήθως να αναπαραχθούν στο σύνολό τους, αν κάτι πάει στραβά κατά τη διαδικασία. Τα μεγάλα σύνολα δεδομένων από την άλλη πλευρά, επειδή έρχονται σε πολλές μορφές και από

πολλές κατευθύνσεις, είναι αδύνατο να αναπαραχθούν από την αρχή, αν κάτι πάει στραβά. Συνήθως το καλύτερο που μπορεί να γίνει είναι να εντοπιστούν τα τμήματα του έργου που είναι προβληματικά, και να κρατηθούν υπόψη, ενώ θα γίνεται εργασία γύρω από αυτά.

ο **Stakes** - Στα Μικρά Δεδομένα, αν τα πράγματα πάνε στραβά, το κόστος μπορεί να είναι περιορισμένο. Στα έργα με Μεγάλα Δεδομένα το κόστος μπορεί να είναι πολύ μεγάλο και αν χαθούν δεδομένα ή υπάρχει αλλοίωση τότε μπορεί να καταδικαστεί το έργο και ακόμα και η ίδια η ύπαρξη του οργανισμού.

ο **Ενδοσκόπηση** - Ενδοσκόπηση σημαίνει ότι τα ίδια τα δεδομένα περιγράφονται με έναν σημαντικό τρόπο. Για τα Μικρά Δεδομένα αυτό θα μπορούσε να περιγραφεί με μία τριάδα που αποτελείται από το όνομα, τον τύπο και την αξία. Επιπλέον, στα Μικρά Δεδομένα, όλα είναι καλά οργανωμένα και είναι συνήθως σαφές για ποιο λόγο χρησιμοποιούνται. Στα Μεγάλα Δεδομένα, επειδή τα πράγματα μπορεί να είναι πιο περίπλοκα, γιατί μπορεί να χρησιμοποιούνται πολλά αρχεία και πολλές μορφές, τα έργα μπορεί να καταλήξουν με δεδομένα που είναι απροσδιόριστα και χωρίς νόημα (αυτό θα μπορούσε να θέσει σε κίνδυνο την χρησιμότητα των Μεγάλων Δεδομένων).

ο **Ανάλυση** - Τα Μικρά Δεδομένα είναι δυνατό να αναλυθούν όλα μία φορά, με μια ενιαία διαδικασία ή σε ένα αρχείο υπολογιστή. Με τα Μεγάλα Δεδομένα τα πράγματα έχουν τεράστιες διαστάσεις και είναι διασκορπισμένα σε διαφορετικά αρχεία και διακομιστές (servers). Έτσι, μπορεί να χρειαστεί να γίνει εξαγωγή, κριτική, ομαλοποίηση, μετατροπή, και άλλα, με ένα υποσύνολο δεδομένων κάθε φορά προκειμένου να καταστούν πιο διαχειρίσιμα. Στο τέλος όλα τα αποτελέσματα συνδυάζονται και δημιουργούνται οι εκθέσεις.

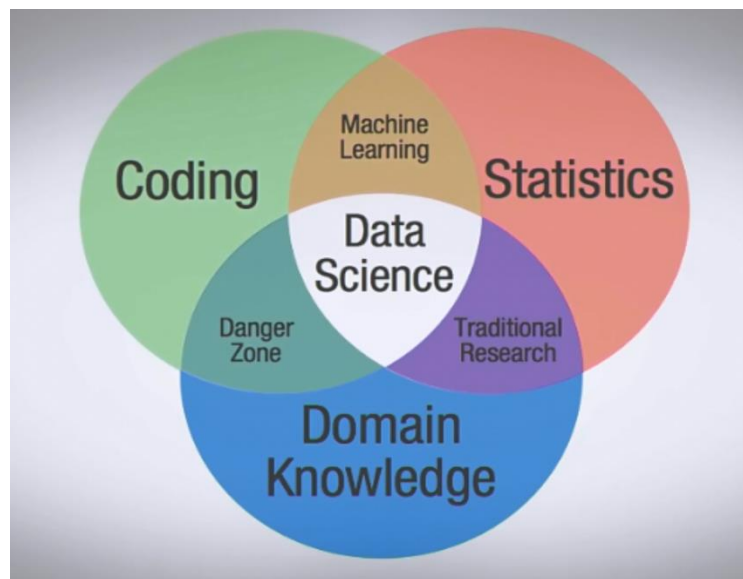
Κλείνοντας υπάρχουν περισσότερες πτυχές των μεγάλων στοιχείων εκτός από τον όγκο, την ταχύτητα, και την ποικιλία που κάνουν τα πράγματα πιο περίπλοκα από ό, τι με τα μικρά στοιχεία.

## 2.6 Τα πρόσωπα της επιστήμης των δεδομένων

Όταν οι άνθρωποι μιλούν για Μεγάλα Δεδομένα, μπορούν επίσης να μιλήσουν για την επιστήμη των δεδομένων και τους επιστήμονες των δεδομένων. Καθώς ο ορισμός των

Μεγάλων Δεδομένων συζητείται ακόμα, το ίδιο ισχύει και για τον ορισμό της επιστήμης των δεδομένων. Για κάποιους είναι εύκολο να τα πουν στατιστικά στοιχεία, αλλά για άλλους η επιστήμη των δεδομένων είναι ένας ξεχωριστός τομέας, δεδομένου ότι έχει διαφορετική εκπαίδευση, εργαλεία και διαφορετικούς στόχους από ό, τι ένα τυπικό στατιστικό στοιχείο.

Ακολουθεί το διάγραμμα της επιστήμης των δεδομένων που δημιουργήθηκε από τον Drew Conway το 2010.



Σχήμα 1 – Επιστήμη των δεδομένων – Διάγραμμα γεγονότων

Με βάση το διάγραμμα η επιστήμη των δεδομένων περιλαμβάνει έναν συνδυασμό τριών διαφορετικών δεξιοτήτων, συμπεριλαμβανομένων της στατιστικής, της γνώσης πεδίου, και της κωδικοποίησης.

## 2.7 Τύποι και ικανότητες της επιστήμης των δεδομένων

Όταν οι άνθρωποι μιλούν για Μεγάλα Δεδομένα είναι εύκολο να καταλάβουμε ότι οι επιστήμονες των δεδομένων δεν είναι μόνο άνθρωποι με εμπειρία στον τομέα ή στατιστικοί εμπειρογνώμονες ή προγραμματιστές. Όπως και με κάθε άλλο τομέα υπάρχουν πολλές άλλες δεξιότητες που εμπλέκονται με την επιστήμη των δεδομένων.

Μια έκθεση με το όνομα "Αναλύοντας τους Αναλυτές - μια εσωτερική έρευνα των επιστημόνων δεδομένων και του έργου τους" από τους Harlan Harris, Sean Murphy και Marck Vaisman δημοσιεύθηκε το 2013. Η έρευνα έγινε σε 250 επαγγελματίες της επιστήμης των δεδομένων, οι οποίοι ερωτήθηκαν για το πώς προσδιορίζουν τον εαυτό τους και τι δεξιότητες χρησιμοποιούν που είναι σχετικές με την επιστήμη των δεδομένων. Αυτό που διαπιστώθηκε ήταν ότι υπάρχει ένα υψηλό επίπεδο ετερογένειας μεταξύ των ανθρώπων των Μεγάλων Δεδομένων.

Σύμφωνα με την έρευνα, υπάρχουν έντεκα πιθανές επαγγελματικές ταυτότητες, όπως φαίνεται παρακάτω:

<b>Data Developer</b>	Developer	Engineer	
<b>Data Researcher</b>	Researcher	Scientist	Statistician
<b>Data Creative</b>	Jack of All Trades	Artist	Hacker
<b>Data Businessperson</b>	Leader	Businessperson	Entrepreneur

Σχήμα 2 – Επαγγελματικές ταυτότητες

Υπάρχουν τέσσερις βασικές κατηγορίες:

ο Developer Δεδομένων - Προγραμματιστής / Μηχανικός

ο Ερευνητής Δεδομένων - Ερευνητές / επιστήμονες / στατιστικοί

ο Άνθρωποι του Δημιουργικού - Jack of All Trades, καλλιτέχνες, χάκερ

ο Ο επαγγελματίας των δεδομένων - Ηγέτης / Επαγγελματίας / Επιχειρηματίας

Επίσης, με βάση την έρευνα, το παρακάτω σχήμα παρουσιάζει τις σημαντικότερες δεξιότητες που απαιτούνται για την επιστήμη των δεδομένων και οι οποίες κατηγοριοποιούνται σε πέντε συμπλέγματα δεξιοτήτων:

Business	ML/BigData	Math/OR	Programming	Statistics
Product Development	Unstructured Data	Optimization	System Administration	Visualization
Business	Structured Data	Math	Back-End Programming	Temporal Statistics
	Machine Learning	Graphical Models	Front-End Programming	Surveys and Marketing
	Big and Distributed Data	Bayesian/Monte Carlo Statistics		Spatial Statistics
		Algorithms		Science
		Simulation		Data Manipulation
				Classical Statistics

Σχήμα 3 – Δεξιότητες της Επιστήμης των Δεδομένων

ο Ο Επιχειρηματίας Δεδομένων - είναι ένα πρόσωπο που μπορεί να εμπλέκεται σε όλους τους τομείς, αλλά πρωταρχικός ρόλος του είναι να είναι ένα επιχειρησιακό πρόσωπο, ηγέτης και επιχειρηματίας

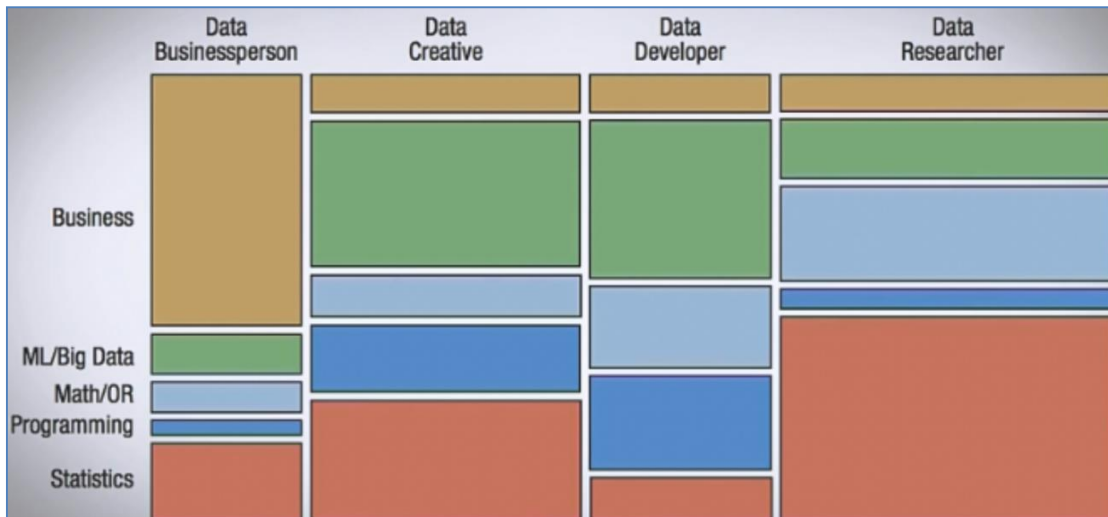
ο Ο άνθρωπος του δημιουργικού - Είναι το πρόσωπο των οποίων οι δεξιότητες είναι οι πιο ομοιόμορφα κατανεμημένες

ο Ο Developer Δεδομένων - Εστιάζει κυρίως σε Μεγάλα Δεδομένα, Μαθηματικά και Προγραμματισμό

ο Ο Ερευνητής Δεδομένων - επικεντρώνεται κυρίως στην στατιστική ανάλυση

Το πιο προφανές είναι ότι κανένας δεν είναι ίδιος με τον άλλο. Κάθε ομάδα έχει κάποια επιδεξιότητα σε κάθε τομέα, αλλά η κατανομή είναι δραματικά διαφορετική. Αυτό που είναι σαφές είναι ότι υπάρχουν περιθώρια για ουσιαστική μεταβολή και προσωπικό ενδιαφέρον σε ομάδες ικανοτήτων στην επιστήμη των δεδομένων και κατ'επέκταση και των Μεγάλων Δεδομένων. Οι ταυτότητες και δεξιότητες φαίνονται στο παρακάτω σχήμα:





Σχήμα 4 – Ταυτότητες και Δεξιότητες

## 2.8 Επιστήμη δεδομένων χωρίς Μεγάλα Δεδομένα [29]

1

Υποστηρίζεται ότι τα Μεγάλα Δεδομένα απαιτούν όλα τα V, Όγκο, Ταχύτητα, και Ποικιλία (Volume, Velocity, and Variety) ταυτόχρονα. Υπάρχουν όμως και περιπτώσεις που η επιστήμη των δεδομένων μπορεί να χρησιμοποιηθεί με τη χρήση ενός V τη φορά, σε μία χρονική στιγμή.

- Επιστήμη δεδομένων με Όγκο

Έτσι, θα μπορούσε να είναι ένα έργο με έναν πολύ μεγάλο όγκο δεδομένων, αλλά χωρίς την ταχύτητα ή την ποικιλία. Αυτό θα σήμαινε ένα πολύ μεγάλο και στατικό σύνολο δεδομένων με μια συνεπή μορφή. Τα δεδομένα θα πρέπει να είναι δομημένα, έτσι για παράδειγμα ελεύθερο κείμενο δεν μπορεί να χρησιμοποιηθεί. Για παράδειγμα, τα δεδομένα γενετικής έχουν τεράστιο όγκο, αλλά έχουν μια πολύ καλά κατανοητή δομή. Ένα άλλο καλό παράδειγμα είναι οι περιπτώσεις της άντλησης δεδομένων και των προγνωστικών αναλυτικών μοντέλων. Στην τελευταία περίπτωση, μπορούν να γίνουν προβλέψεις για ένα αποτέλεσμα όπως αν ένα άτομο θα κάνει κλικ σε μια διαφήμιση ή μια ιστοσελίδα. Στην περίπτωση αυτή, μπορεί να δημιουργηθεί ένα σύνολο δεδομένων με χιλιάδες μεταβλητές ή ακόμη και δισεκατομμύρια περιπτώσεις, αλλά όλα τα δεδομένα θα έχουν μία συνεπή μορφή. Το μέγεθος των στοιχείων καθιστά πολλές κοινές προσεγγίσεις αδύνατες, έτσι οι ικανότητες προγραμματισμού για

<sup>1</sup> Βιβλιογραφία για τις ενότητες 2.8, 2.9, 2.10: [30, 31, 32, 33]

πολλούς επιστήμονες των δεδομένων είναι απαραίτητες, μαζί με τη γνώση του τομέα και της στατιστικής.

- Επιστήμη δεδομένων με Ταχύτητα

Το επόμενο παράδειγμα είναι η επιστήμη των δεδομένων με Ταχύτητα χωρίς Όγκο ή Ποικιλία. Ένα πολύ καλό παράδειγμα είναι η συνεχής ροή δεδομένων με μια συνεπή δομή. Η συνεχής ροή δεδομένων (data streaming) είναι μια τεχνική όπου μόνο ένα μικρό τμήμα των δεδομένων διατηρείται. Ένα ενδιαφέρον παράδειγμα είναι τα συστήματα ανίχνευσης σεισμών της γεωλογικής έρευνας των Ηνωμένων Πολιτειών. Το σύστημα ελέγχει αν γίνονται σεισμοί ή πρόκειται να συμβούν. Στην περίπτωση αυτή δεν είναι απαραίτητο να κρατηθούν όλα τα δεδομένα, εάν η μόνη αναγκαία πράξη είναι να πυροδοτηθεί ένα συμβάν. Έτσι, δεν είναι αναγκαίο να κρατηθούν τα δεδομένα, πράγμα που κάνει τον όγκο σχετικά μικρό, και μπορεί να έχουν μια πολύ συνεπή δομή με αποτέλεσμα να έχουν και μικρή ποικιλία. Η υπόθεση αυτή θα μπορούσε επίσης να κληθεί άντληση ροής δεδομένων. Και ένα πιθανό παράδειγμα θα μπορούσε να είναι η ταξινόμηση των δεδομένων των αισθητήρων σε πραγματικό χρόνο.

- Επιστήμη των δεδομένων και ποικιλία

Τέλος, υπάρχει η περίπτωση της επιστήμης των δεδομένων με την ποικιλία. Εδώ τα δεδομένα μπορούν να υπάρχουν σε πολλές διαφορετικές μορφές, αλλά χωρίς μεγάλο όγκο, και ταχύτητα. Αυτή η περίπτωση μπορεί να είναι πιο περίπλοκη, αλλά με μικρότερα και σχετικά στατικά σύνολα δεδομένων. Μερικά παραδείγματα περιλαμβάνουν την αναγνώριση προσώπου στις προσωπικές συλλογές φωτογραφιών όπου δεν υπάρχουν πολλές φωτογραφίες, αλλά υπάρχει μια μεγάλη ποικιλία σε οπτικά δεδομένα.

- Επιστήμη δεδομένων με περισσότερα από ένα V

Όλα τα παραδείγματα που αναφέρονται παραπάνω χρησιμοποιούν την επιστήμη των δεδομένων και απαιτούν στατιστική γνώση, εμπειρία στον τομέα, ή ακόμη και γνώσεις προγραμματισμού, ενώ υπάρχει ένα V τη φορά. Παρακάτω παρατίθενται παραδείγματα όπου περιλαμβάνονται περισσότερα από ένα V.

## Όγκος και ταχύτητα

Υπάρχουν επίσης περιπτώσεις όπου περισσότερα από ένα V μπορούν να συνυπάρχουν. Για παράδειγμα, υπάρχουν περιπτώσεις που περιλαμβάνουν όγκο και ταχύτητα, αλλά όχι πολλή ποικιλία. Για παράδειγμα, όταν πολλά δεδομένα εγγράφονται πολύ γρήγορα και είναι στην ίδια μορφή. Ενδιαφέροντα παραδείγματα περιλαμβάνουν στοιχεία για τη χρηματιστηριακή αγορά και δεδομένα από τους τζετ κινητήρες αεροπλάνων. Ένα ενδιαφέρον γεγονός είναι ότι οι μηχανές τζετ διαθέτουν αισθητήρες που μπορούν να παράγουν 20TB πληροφοριών κάθε ώρα, που είναι μια τεράστια πηγή πληροφοριών. Αυτός ο τύπος δεδομένων είναι ζωτικής σημασίας να διατηρείται, επειδή η αποτυχία ενός κινητήρα τζετ είναι ένα εξαιρετικά σημαντικό πράγμα. Μέσω αυτών των δεδομένων οι επιστήμονες μπορούν να βρουν τι προκάλεσε την αποτυχία.

## Ταχύτητα και Ποικιλία

Μια άλλη προοπτική είναι η επιστήμη των δεδομένων που εφαρμόζεται σε δεδομένα τα οποία έχουν ταχύτητα, πολλή μεγάλη ποικιλία αλλά όχι μεγάλο όγκο. Αυτό μπορεί να συμβεί με τη συνεχή ροή δεδομένων (data streaming) όπου δεν είναι απαραίτητο να κρατηθούν όλα τα δεδομένα. Ένα ενδιαφέρον παράδειγμα είναι το βίντεο επιτήρησης. Σύμφωνα με την πρόσφατη έκθεση της International Data Corporations, «Το ψηφιακό σύμπαν το 2020», το ήμισυ των παγκόσμιων Μεγάλων Δεδομένων - το πολύτιμο θέμα για ανάλυση στο ψηφιακό σύμπαν - ήταν το βίντεο επιτήρησης το 2012, και το ποσοστό αυτό αναμένεται να αυξηθεί στο 65 τοις εκατό μέχρι το 2015. Ένας λόγος που θα συμβεί αυτό είναι γιατί τα δεδομένα επιτήρησης θα αναβαθμιστούν σε δεδομένα υψηλής ευκρίνειας. Η πληροφόρηση που έρχεται είναι πολύ γρήγορη, επεξεργάζεται πολύ γρήγορα και έχει μεγάλη ποικιλία, διότι είναι οπτικά δεδομένα. Στην περίπτωση αυτή, ο στόχος είναι να αναλύονται τα δεδομένα πολύ γρήγορα και να ενεργοποιούνται ειδικές δράσεις, όταν είναι απαραίτητο.

## Όγκος και Ποικιλία

Η τελευταία περίπτωση είναι ο όγκος και η ποικιλία, αλλά χωρίς ταχύτητα. Αυτό περιλαμβάνει ιστορικά δεδομένα που έχουν πολλαπλές μορφές ή περιλαμβάνουν οπτικά δεδομένα. Ένα πολύ καλό παράδειγμα είναι το Google Books. Η Google έχει πάνω από τριάντα εκατομμύρια βιβλία που έχουν σαρωθεί και ψηφιοποιηθεί. Η πληροφόρηση αυτή είναι αρκετά μεγάλη σε όγκο και περίπλοκη. Παρόμοια παραδείγματα περιλαμβάνουν το Twitter όπου κάθε tweet που έχει γραφτεί ποτέ

αποθηκεύεται. Αυτό είναι ένα τεράστιο μέγεθος πληροφοριών, διότι το κείμενο είναι ένα σύνθετο δεδομένο, αλλά επειδή δεν ενημερώνεται συνεχώς δεν υφίσταται η ταχύτητα.

- Συμπεράσματα

Τα παραδείγματα δείχνουν ότι, παρά την ισχυρή συσχέτιση μεταξύ των Μεγάλων Δεδομένων και της επιστήμης των δεδομένων, οι δεξιότητες της επιστήμης των δεδομένων (στατιστικές γνώσεις, εξειδίκευση στον τομέα, και δεξιότητες προγραμματισμού) εφαρμόζονται ακόμη και όταν οι τρεις πτυχές των Μεγάλων Δεδομένων δεν συνυπάρχουν την ίδια στιγμή.

## 2.9 Μεγάλα Δεδομένα χωρίς την επιστήμη των δεδομένων

Τα Μεγάλα Δεδομένα περιλαμβάνουν ασυνήθιστο όγκο, ταχύτητα και ποικιλία. Η επιστήμη των δεδομένων περιλαμβάνει δεξιότητες στατιστικής, γνώση του αντικειμένου, καθώς και ικανότητα προγραμματισμού.

Το θέμα είναι εάν μπορούμε να έχουμε Μεγάλα Δεδομένα με ένα υποσύνολο των προαναφερθέντων δεξιοτήτων.

### Στατιστική και Κωδικοποίηση

Για παράδειγμα, αν χρησιμοποιηθεί Στατιστική και Κωδικοποίηση, μπορεί να χρησιμοποιηθεί Μηχανική Μάθηση για την εκπλήρωση των έργων. Η Μηχανική Μάθηση είναι ένας πολύ σημαντικός τομέας της επιστήμης των δεδομένων, όπου ένα πρόγραμμα υπολογιστή μαθαίνει να προσαρμόζεται στις νέες πληροφορίες που εισέρχονται. Τα δύο πιο γνωστά παραδείγματα είναι τα φίλτρα spam όπου το πρόγραμμα υπολογιστή μαθαίνει να προσδιορίζει αν ένα συγκεκριμένο είδος ηλεκτρονικού ταχυδρομείου είναι spam με βάση τις απαντήσεις των εκατομμυρίων χρηστών (για παράδειγμα, στο Gmail). Το δεύτερο παράδειγμα είναι η αναγνώριση προσώπου σε φωτογραφίες, όπου το πρόγραμμα υπολογιστή «μαθαίνει» ποιο πρόσωπο ανήκει σε ποιον. Έτσι, η Μηχανική Μάθηση είναι ένα καλό παράδειγμα ενασχόλησης με Μεγάλα Δεδομένα, διότι μπορεί να συνυπάρχουν ο Όγκος, η Ταχύτητα, αλλά και η Ποικιλία. Ωστόσο, δεν χρειάζεται η γνώση του αντικειμένου, διότι το λογισμικό του υπολογιστή μπορεί να επεξεργαστεί και να ελέγξει τα αποτελέσματα.

## Κωδικοποίηση και Γνώση Αντικειμένου

Μια άλλη περίπτωση είναι ο συνδυασμός της κωδικοποίησης και της γνώσης του αντικειμένου χωρίς τη γνώση της στατιστικής. Υπάρχουν πολλά επικίνδυνα ζητήματα που μπορούν να δημιουργηθούν εδώ, αλλά υπάρχουν πολλά παραδείγματα αυτής της περίπτωσης όπου δεν εμπλέκονται τα στατιστικά στοιχεία. Μερικά παραδείγματα περιλαμβάνουν μέτρηση λέξεων και ανάλυση της φυσικής γλώσσας. Το πιο κοινό εργαλείο που χρησιμοποιεί αυτόν τον συνδυασμό είναι το «Εργαλείο Φυσικής Γλώσσας», το οποίο χρησιμοποιείται από την Python και επιτρέπει στους ανθρώπους να κάνουν πολλές εργασίες. Για παράδειγμα, δημιουργήθηκε ένα εργαλείο το οποίο μετρά τον αριθμό των φορών που μια λέξη εμφανίζεται σε ένα κείμενο, πράγμα που δεν απαιτεί καμία στατιστική γνώση.

## Περιορισμοί

Χωρίς δεξιότητες προγραμματισμού δεν είναι δυνατόν να ασχοληθεί κανείς με τον όγκο, την ταχύτητα ή την ποικιλία που χαρακτηρίζει τα Μεγάλα Δεδομένα. Τεράστια πράγματα μπορούν να επιτευχθούν με απλή έρευνα, αλλά τα Μεγάλα Δεδομένα δεν είναι ανάμεσα σε αυτά.

Επιπλέον, δεν είναι δυνατόν να χρησιμοποιηθούν Μεγάλα Δεδομένα, αν υπάρχει μόνο μία από τις τρεις ικανότητες. Στην περίπτωση αυτή θα πρέπει να συνεργαστούν οι άνθρωποι που έχουν αυτές τις ικανότητες. Η συνεργασία είναι το κλειδί για την επιτυχία στην επιστήμη των δεδομένων, από εκεί και πέρα χρειάζεται ένα ευρύ φάσμα δεξιοτήτων που κανείς δεν είναι σε θέση να τις έχει όλες. Έτσι, η επιστήμη των δεδομένων ενθαρρύνει σθεναρά τη συνεργασία με σκοπό τη διεκπεραίωση ορισμένων εργασιών. Επίσης, εάν κοιτάξει κανείς τη σχέση μεταξύ της επιστήμης των δεδομένων και των Μεγάλων Δεδομένων, είναι δυνατόν να έχουμε επιστήμη δεδομένων με ένα ελλιπές σύνολο Μεγάλων Δεδομένων, αλλά είναι πολύ δύσκολο να έχουμε Μεγάλα Δεδομένα χωρίς τις δεξιότητες της επιστήμης των δεδομένων.

## 2.10 Δεδομένα που παράγονται από τον άνθρωπο

Τα Μεγάλα Δεδομένα μπορεί να προέρχονται από διάφορες πηγές. Τα δεδομένα μπορούν να παράγονται από ανθρώπους ή μπορεί να παράγονται από μηχανές.

Οι άνθρωποι παράγουν πολλά δεδομένα είτε εσκεμμένα είτε ακούσια. Για παράδειγμα, οι άνθρωποι μπορούν σκόπιμα να δημιουργήσουν δεδομένα που να περιλαμβάνουν φωτογραφίες, βίντεο, ήχο, κείμενο σε ένα κοινωνικό δίκτυο, likes, αναζητήσεις στο διαδίκτυο, σελιδοδείκτες, ηλεκτρονικό ταχυδρομείο, μηνύματα κειμένου, κλήσεις κινητού τηλεφώνου, κύρια σημεία ηλεκτρονικών βιβλίων, σημειώσεις, αγορές μέσω διαδικτύου, κ.λπ. Όλα τα ανωτέρω στοιχεία είναι δεδομένα που δεν υπάρχουν μέχρι το αντίστοιχο άτομο να τα δημιουργήσει. Είναι στοιχεία/δεδομένα που προκύπτουν από ανθρώπινες πράξεις.

Το ενδιαφέρον πράγμα σχετικά με αυτές τις πληροφορίες είναι ότι συνοδεύονται από ένα είδος πληροφοριών που ονομάζονται μεταδεδομένα. Τα μεταδεδομένα είναι δεδομένα για τα δεδομένα ή ονομάζονται επίσης δευτερεύοντα δεδομένα που παράγονται από τα δεδομένα των ανθρώπων. Η ιδέα είναι ότι πρόκειται για δεδομένα που συνοδεύουν αυτά που κάνει ένα άτομο, αλλά το άτομο δεν τα γνωρίζει. Επίσης, τα παραγόμενα μεταδεδομένα μπορεί να είναι τεράστια και μερικές φορές μεγαλύτερα από ό, τι τα δεδομένα που παράγονται από τον χρήστη, και πιο σημαντικά για τον κόσμο των Μεγάλων Δεδομένων, γιατί δημιουργούνται από τον υπολογιστή και είναι αναγνώσιμα και αναζητήσιμα. Παραδείγματα πηγών μεταδεδομένων περιλαμβάνουν φωτογραφίες, κινητό τηλέφωνο και ηλεκτρονικό ταχυδρομείο.

Το MIT δημιούργησε μια διαδικτυακή εφαρμογή που ονομάζεται Βύθιση (Immersion) η οποία επιτρέπει στο χρήστη να κάνει μια γρήγορη ανάλυση του δικού του κοινωνικού δικτύου μέσω λογαριασμού ηλεκτρονικού ταχυδρομείου. Τα αποτελέσματα μπορούν να δείξουν, για παράδειγμα, όλα τα άτομα που ανήκουν σε μία συγκεκριμένη ομάδα. Ένα άλλο παράδειγμα περιλαμβάνει το Twitter. Κάθε tweet είναι μία πολύ μικρή πληροφορία και περιορίζεται σε εκατόν σαράντα χαρακτήρες. Από άποψη έρευνας, υπάρχει ένα τεράστιο ποσό μεταδεδομένων που οι εταιρείες μπορούν να βρουν σε κάθε tweet. Το Twitter έχει ένα πολύ πλούσιο σύνολο δεδομένων για τους ανθρώπους που κάνουν έρευνα μάρκετινγκ ή έρευνα κοινωνικής δικτύωσης. Κλείνοντας, το πιο ενδιαφέρον είναι ότι τα μεταδεδομένα δεν χρειάζεται να υποβληθούν σε επεξεργασία, επειδή είναι ήδη αναγνώσιμα από υπολογιστή, μπορούν να αναζητηθούν, και μπορούν

να ανακτηθούν αμέσως προκειμένου να επιτευχθούν οι στόχοι της ανάλυσης των Μεγάλων Δεδομένων.

## 2.11 Δεδομένα που παράγονται από μηχανές

Εκτιμάται ότι το 90% των δεδομένων στον κόσμο δεν θα ειδωθεί ποτέ από τα ανθρώπινα μάτια. Μεγάλο μέρος αυτών των άορατων δεδομένων ονομάζονται στοιχεία από μηχανή προς μηχανή. Οι μηχανές μπορεί να μιλούν μεταξύ τους και να μην ακούγονται από τους ανθρώπους.

Ορισμένες από τις πηγές των δεδομένων που δημιουργούνται από μηχανές, περιλαμβάνουν:

- Κινητά τηλέφωνα που συνδέονται με πύργους
- Δορυφορικό ραδιόφωνο και GPS σύνδεση
- Αναγνώσεις RFID
- Ενδείξεις από ιατρικές συσκευές

## 2.12 Το Διαδίκτυο των Πραγμάτων - Internet of Things

Ίσως το πιο ενδιαφέρον μέρος των δεδομένων από μηχανή προς μηχανή είναι το «Διαδίκτυο των πραγμάτων (ΔΤΠ)». Εκτιμάται ότι μέχρι το 2020 τριάντα δισεκατομμύρια μοναδικά αναγνωρίσιμες συσκευές θα συνδεθούν στο Διαδίκτυο. Έτσι ουσιαστικά τα πάντα θα έχουν ένα μικροτσίπ το οποίο θα συνδέεται με το διαδίκτυο, θα μιλάμε ο ένας στον άλλον και θα μοιραζόμαστε πληροφορίες.

Μια άλλη τάση είναι αυτή που χρησιμοποιούνται έξυπνοι αισθητήρες στα έξυπνα σπίτια (κλιματιστικά, φώτα, έλεγχος θερμοκρασίας), καθώς και στις πόλεις (έλεγχος του ηλεκτρικού δικτύου, συντονισμός της κίνησης και των επιχειρήσεων κοινής ωφέλειας). Το πλεονέκτημα της χρήσης έξυπνων αισθητήρων είναι ότι μπορεί να συμβάλει στη βελτίωση της αποτελεσματικότητας, στην κατανάλωση, και σε όλες τις υπηρεσίες. Όλο αυτό μπορεί να επιτευχθεί με τη χρήση μικρών συσκευών που θα επικοινωνούν μεταξύ τους στο διαδίκτυο των πραγμάτων (επικοινωνούν απευθείας και χωρίς ενδιάμεσο).

Μερικές από τις χρήσεις αυτού, θα μπορούσαν να περιλαμβάνουν:

- Παρακολούθηση γραμμών παραγωγής
- Έξυπνοι μετρητές για συστήματα κοινής ωφελείας, μετρητές νερού, ρεύματος
- Θερμοστάτες, λαμπτήρες
- Παρακολούθηση του περιβάλλοντος
- Διαχείριση υποδομών
- Βιομηχανικές εφαρμογές
- Διαχείριση της ενέργειας
- Συστήματα για ιατρική και υγεία
- Αυτοματισμός κατασκευών και σπιτιού

Ο κανόνας είναι ότι οτιδήποτε μηχανικό μπορεί να συνδεθεί και να επικοινωνήσει μέσω του διαδικτύου των πραγμάτων. Αυτές οι μηχανικές συσκευές μπορούν να δημιουργήσουν ένα τεράστιο όγκο δεδομένων, προκειμένου να επικοινωνήσουν η μία με την άλλη και να συντονίσουν τις δραστηριότητές τους.

### 2.13 Διαφορές μεταξύ των δεδομένων που παράγονται από ανθρώπους και των δεδομένων που παράγονται από μηχανές

Η πιο εμφανής διαφορά είναι ότι όλα τα δεδομένα που προκύπτουν από μηχανές είναι αναγνωρίσιμα από μηχανές (μπορούν να αναζητηθούν πολύ εύκολα, να διαβαστούν). Έχουν μεγάλο όγκο και ταχύτητα, αλλά χαμηλή ποικιλία, πράγμα που καθιστά ευκολότερο για τις μηχανές να επεξεργαστούν αυτά τα δεδομένα.

### 2.14 Δομημένα δεδομένα

Τα δεδομένα που θεωρούνται δομημένα είναι τοποθετημένα σε ένα αρχείο με σταθερά πεδία ή μεταβλητές. Το πιο γνωστό είδος δομημένων δεδομένων είναι το υπολογιστικό φύλλο. Στον επιχειρηματικό κόσμο τα μεγάλα σύνολα δεδομένων αποθηκεύονται σε βάσεις δεδομένων, οι οποίες έχουν αρκετά μεγάλη χωρητικότητα, ευελιξία και



συνέπεια. Μια πρόσφατη έρευνα έδειξε ότι σχεδόν το 80% χρησιμοποιούν κάποια μορφή βάσης δεδομένων συμπεριλαμβανομένων των MySQL, Microsoft SQL ή Oracle.

## 2.15 Αδόμητα δεδομένα

Εκτός από τα δομημένα δεδομένα, υπάρχουν και τα αδόμητα δεδομένα. Ένα παράδειγμα είναι ένα απλό αρχείο κειμένου. Το θέμα με τα αδόμητα δεδομένα είναι ότι δεν είναι εύκολο για ένα πρόγραμμα υπολογιστή να τα διαβάσει, να τα επεξεργαστεί, να τα ταξινομήσει, να τους αλλάξει σειρά, να μετρήσει τις τιμές και να προσθέσει περισσότερες παρατηρήσεις.

Έτσι τα αδόμητα δεδομένα:

- Είναι δεδομένα που δεν είναι σε σταθερά πεδία
- Περιλαμβάνουν κείμενα, παρουσιάσεις, εικόνες, κλπ
- Είναι η πλειοψηφία των δεδομένων στις επιχειρήσεις
- Για να μετατραπούν, είναι αναγκαία προγράμματα άντλησης κειμένου
- Η διαδικασία ενασχόλησης με αυτά είναι δύσκολη και χρονοβόρα

## 2.16 Ημι-δομημένα δεδομένα

Ημι-δομημένα δεδομένα, είναι δεδομένα που δεν είναι σε σταθερά πεδία, ωστόσο τα πεδία αυτά ξεχωρίζουν και τα δεδομένα εξακολουθούν να είναι αναγνωρίσιμα. Δύο κοινές μορφές είναι:

- XML (Extensive Markup Language)
- JSON (JavaScript Object Notation)

Τα ημι-δομημένα και μη-δομημένα δεδομένα συνήθως λειτουργούν με NoSQL βάσεις δεδομένων (και όχι μόνο με SQL). Ο λόγος που χρησιμοποιούνται αυτές οι βάσεις δεδομένων είναι επειδή είναι εξαιρετικά ευέλικτες και μπορούν να χειριστούν ένα ευρύ φάσμα μορφών δεδομένων. Μερικά ενδιαφέροντα στοιχεία σχετικά με NoSQL βάσεις δεδομένων είναι τα εξής:

- Χρήση ημι-δομημένης μορφής
- Ευέλικτη δομή και συχνά πιο γρήγορη
- Ένα τεράστιο μέγεθος των μεγάλων δεδομένων αποθηκεύεται σε βάσεις δεδομένων NoSQL
- Ένα μειονέκτημα είναι ότι δεν έχει τυποποιημένη γλώσσα αναζητήσεων

## 2.17 Προκλήσεις σχετικά με την ποιότητα των δεδομένων

Σε οποιοδήποτε έργο που περιλαμβάνει δεδομένα, η ποιότητα των δεδομένων διαδραματίζει έναν κρίσιμο ρόλο. Στα Μεγάλα Δεδομένα, η ποιότητα είναι ακόμα πιο δύσκολο να επιτευχθεί, επειδή συχνά προέρχονται από δεδομένα εκτός του έργου και θα χρησιμοποιηθούν και μετά το πέρας του αυτού. Οι ερευνητές δεν έχουν τον πλήρη έλεγχο των δεδομένων όπως έχουν στα έργα με Μικρά Δεδομένα. Επίσης, πολλά από τα δεδομένα που χρησιμοποιούνται σε ένα έργο Μεγάλων Δεδομένων μπορούν να εξαχθούν από υπολογιστικά φύλλα. Οι ερευνητές έχουν βρει ότι σχεδόν το 95% των λογιστικών φύλλων που εξετάστηκαν έχουν σφάλματα. Αυτό θα μπορούσε να είναι ένα ζήτημα δεδομένου ότι, όταν ένα πρόγραμμα χρησιμοποιεί λάθος δεδομένα εισόδου, τότε πιθανότατα, τα αποτελέσματα θα είναι επίσης λανθασμένα. Έτσι, είναι απαραίτητο τα δεδομένα να είναι σωστά από την αρχή, ώστε τα αποτελέσματα να είναι έγκυρα.

Ορισμένες από τις προκλήσεις που πρέπει να εξεταστούν όσον αφορά την ποιότητα των δεδομένων είναι:

- Ελλιπή ή κατεστραμμένα δεδομένα - Αυτό θα μπορούσε να οδηγήσει σε μηδενικούς δείκτες, όπου οι υπολογιστές θα έψαχναν για κάτι που δεν υπάρχει
- Διπλές εγγραφές - πράγμα που σημαίνει ότι τα ίδια στοιχεία εμφανίζονται πολλές φορές
- Τυπογραφικά λάθη - Λάθος εισαγωγή δεδομένων θα μπορούσαν να οδηγήσουν σε μεγάλα σφάλματα
- Τα δεδομένα από τα οποία λείπει το πλαίσιο – η έλλειψη πλαισίου θα μπορούσε να οδηγήσει σε δαπανηρά και χρονοβόρα σφάλματα

Όλα αυτά τα θέματα θα μπορούσαν να επηρεάσουν ένα μικρό έργο σε έκταση, αλλά θα μπορούσαν να επηρεάσουν ακόμα περισσότερο ένα έργο Μεγάλων Δεδομένων.

Προβλήματα με ένα έργο Μεγάλων Δεδομένων:

- Κανονικοί μέθοδοι ελέγχου δεν μπορούν να χρησιμοποιηθούν
- Εάν τα δεδομένα υπάρχουν ήδη σε μια βάση δεδομένων NoSQL, μπορεί να μην πρέπει να μετατραπούν
- Αν τα δεδομένα δεν μετατραπούν, τότε δεν μπορούν να χρησιμοποιηθούν

Υπάρχουν επίσης ορισμένες ειδικές προκλήσεις που αφορούν τα Μεγάλα Δεδομένα:

- Οι αρμοδιότητες θα πρέπει να μοιραστούν σε έναν μεγάλο αριθμό ατόμων, έτσι ώστε διαφορετικοί άνθρωποι να μπορούν να προετοιμάσουν, να αναλύσουν και να εφαρμόσουν τα δεδομένα
- Υπάρχει μεγαλύτερη ανάγκη για την εξασφάλιση της ποιότητας
- Υπάρχει μεγαλύτερη ανάγκη για τον προσδιορισμό της έννοιας του έργου

## 3 Αναλυτικά Μοντέλα για Μεγάλα Δεδομένα

### 3.1 Εισαγωγή

Η ανάλυση των Μεγάλων Δεδομένων είναι «η συλλογή δεδομένων και η τεχνολογία που δίνει πρόσβαση, ενσωματώνει και δημιουργεί εκθέσεις για όλα τα διαθέσιμα δεδομένα, φιλτράροντας, συσχετίζοντας, παράγοντας ιδέες που δεν μπορούν να επιτευχθούν με τις τεχνολογίες δεδομένων του παρελθόντος» [11].

Επίσης, η ανάλυση των Μεγάλων Δεδομένων θα μπορούσε να χαρακτηριστεί ως η διαδικασία εξέτασης των μεγάλων συνόλων δεδομένων που περιέχουν μια ποικιλία τύπων δεδομένων, ώστε να αποκαλυφθούν κρυμμένα μοτίβα, άγνωστοι συσχετισμοί, τάσεις αγοράς, προτιμήσεις καταναλωτών και άλλες χρήσιμες επιχειρηματικές πληροφορίες. Τα αναλυτικά αποτελέσματα μπορούν να οδηγήσουν σε πιο αποτελεσματικό μάρκετινγκ, νέες ευκαιρίες εσόδων, καλύτερη εξυπηρέτηση πελατών, βελτιωμένη επιχειρησιακή απόδοση, ανταγωνιστικά πλεονεκτήματα έναντι των ανταγωνιστών και άλλα επιχειρηματικά οφέλη.

Η ανάλυση των Μεγάλων Δεδομένων είναι ένα εξελισσόμενο φαινόμενο με σημαντική ανταπόκριση τόσο στους επαγγελματίες και όσο και στους ακαδημαϊκούς. Μερικοί ερευνητές έχουν πει ότι τα Μεγάλα Δεδομένα είναι το «επόμενο μεγάλο πράγμα στην καινοτομία» [13], "το τέταρτο παράδειγμα της επιστήμης» [18], καθώς επίσης και το «επόμενο σύνορο για την καινοτομία, τον ανταγωνισμό και την παραγωγικότητα» [14].

Τα Μεγάλα Δεδομένα, αν και είναι ακόμη μία αναδυόμενη έννοια, αντιμετωπίζονται ταυτόχρονα σαν μία πραγματικότητα. Ωστόσο, τι μπορεί πραγματικά να σημαίνουν για έναν οργανισμό από τεχνολογικής και οργανωτικής άποψης στη διαχείρισης των πόρων; Τα Μεγάλα Δεδομένα είναι μια επανάσταση και έχουν την ικανότητα να αλλάξουν τον τρόπο με τον οποίο οι οργανισμοί λειτουργούν [15]. Στην εποχή των Μεγάλων Δεδομένων, οι οργανισμοί έχουν στη διάθεσή τους τεράστια ποσότητα πληροφοριών που μπορούν να τους παρέχουν πολύτιμες ιδέες, περισσότερο για το τι, παρά για το γιατί. Με τη χρήση κατάλληλης τεχνολογίας και πόρων μπορεί να αποκτηθεί ένα σημαντικό ανταγωνιστικό πλεονέκτημα.

Σύμφωνα με το International Data Corporation (IDC) - 2011, οι τεχνολογίες των Μεγάλων Δεδομένων περιγράφουν μια νέα γενιά τεχνολογιών και αρχιτεκτονιών, σχεδιασμένων για να αποσπαστούν οικονομική αξία από πολύ μεγάλο όγκο μιας ευρείας ποικιλίας δεδομένων, επιτρέποντας υψηλή ταχύτητα, ανακάλυψη, ή / και ανάλυση. Τα 7Vs (Volume, Variety, Velocity, Variability, Veracity, Visualization and Value) Όγκος, Ποικιλία, Ταχύτητα, Μεταβλητότητα, Ειλικρίνεια, Απεικόνιση και Αξία των δεδομένων που υπάρχουν στους οργανισμούς, εξακολουθούν να φτάνουν σε υψηλά επίπεδα. Αυτή η αξιοσημείωτη εξέλιξη σημαίνει ότι οι διοικήσεις πρέπει όχι μόνο να αναγνωρίζουν τη σημασία των Μεγάλων Δεδομένων, έτσι ώστε να αποκωδικοποιούν την πληροφορία που έχει ουσιαστική σημασία για αυτούς, αλλά και να κατανοούν τις δυνατότητες της ανάλυσης των Μεγάλων Δεδομένων τόσο στον δημόσιο όσο και στον ιδιωτικό τομέα.

Όπως αναφέρει ο Eric Schmidt (Εκτελεστικός Πρόεδρος, Google) "Από την αυγή του πολιτισμού μέχρι το 2003, η ανθρωπότητα δημιούργησε 5 Exabytes δεδομένων. Τώρα παράγουμε 5 Exabytes κάθε 2 μέρες ... και ο ρυθμός αυτός επιταχύνεται." Αυτή η φράση ακολουθείται από τον Robert J. Moore (RJMetrics Συν-Ιδρυτή και Διευθύνων Σύμβουλο), ο οποίος αναφέρει ότι "23 Exabytes στοιχείων έχουν καταγραφεί και αναπαραχθεί μέχρι το 2002. Τώρα καταγράφεται και μεταφέρεται αυτός ο όγκος πληροφορίας κάθε 7 ημέρες ". Έτσι, με το διαρκώς αυξανόμενο μέγεθος των Μεγάλων Δεδομένων, καθίσταται εξαιρετικά σημαντικό για τους οργανισμούς να βγάλουν νόημα από αυτά τα δεδομένα και πληροφορίες με έγκαιρο και αποτελεσματικό τρόπο - και εδώ είναι όπου η ανάλυση έρχεται να παίξει το ρόλο της.

Μια σειρά από μελέτες έχουν επισημάνει ότι οι οργανισμοί που χρησιμοποιούν επιχειρηματικά αναλυτικά μοντέλα για την αυτοματοποίηση της λήψης αποφάσεων έχουν πιο παραγωγικά αποτελέσματα, όπως βελτιωμένες επιδόσεις, σε βάθος κατανόηση της συμπεριφοράς των καταναλωτών, καλύτερο σχεδιασμό και πρόβλεψη, απολαμβάνουν υψηλότερες αποδόσεις ιδίων κεφαλαίων (ή αξία όσον αφορά στο δημόσιο τομέα), επιτρέποντας στους οργανισμούς να λειτουργούν πιο αποτελεσματικά και να παραμένουν μπροστά από τον ανταγωνισμό [16, 17].

Η αποτελεσματική διαχείριση και αξιοποίηση των δεδομένων είναι πάντα ένα πρόβλημα για τους οργανισμούς. Η κλίμακα και ο σκοπός του μετασχηματισμού των δεδομένων σε αξία έχουν αναγκάσει τους επαγγελματίες και τους ακαδημαϊκούς να διεξάγουν εμπειρικές μελέτες. Ο πλούτος και το τεράστιο μέγεθος των Μεγάλων Δεδομένων δημιουργεί πρωτόγνωρες δυνατότητες έρευνας σε πολλούς τομείς, όπως

φαίνεται και από την εφαρμογή τους στους τομείς της βιομηχανίας, των χρηματοοικονομικών υπηρεσιών, της ασφάλισης, του λιανικού εμπορίου και της υγείας. Στο πλαίσιο της Τεχνολογίας και Διαχείρισης Πόρων, τα Μεγάλα Δεδομένα και τα Αναλυτικά Μοντέλα έχουν τη δυνατότητα να διευκολύνουν την πιο εξελιγμένη λήψη αποφάσεων, με γνώμονα τα δεδομένα, τις επιχειρηματικές διαδικασίες σε πραγματικό χρόνο και νέους τρόπους οργάνωσης και διαχείρισης των δεδομένων.

### 3.2 Στόχοι

Ο πρωταρχικός στόχος της ανάλυσης των μεγάλων δεδομένων είναι να βοηθήσουν τις επιχειρήσεις να λαμβάνουν πιο σωστές επιχειρηματικές αποφάσεις, δίνοντας τη δυνατότητα στους επιστήμονες των δεδομένων, στους μοντελιστές προβλέψεων και άλλους επαγγελματίες της ανάλυσης να αναλύουν μεγάλο όγκο δεδομένων συναλλαγών, καθώς και άλλες μορφές δεδομένων που μπορεί να είναι αναξιοποίητες με τα συμβατικά προγράμματα επιχειρηματικής ευφυΐας (Business Intelligence). Αυτές θα μπορούσαν να περιλαμβάνουν τα αρχεία καταγραφής του διακομιστή διαδικτύου, δεδομένα από κλικς στο διαδίκτυο, περιεχόμενο των μέσων κοινωνικής δικτύωσης, το περιεχόμενο εκθέσεων δραστηριότητας στο κοινωνικό δίκτυο, το κείμενο από μηνύματα ηλεκτρονικού ταχυδρομείου των πελατών, απαντήσεις έρευνας, αναλυτικές εγγραφές κλήσεων κινητής τηλεφωνίας και δεδομένων μηχανής από αισθητήρες που συνδέονται με το Διαδίκτυο των πραγμάτων (Internet of Things). Μερικοί άνθρωποι συνδέουν αποκλειστικά τα Μεγάλα Δεδομένα με ημι-δομημένα και αδόμητα δεδομένα. Ωστόσο, εταιρείες συμβούλων, όπως η Gartner Inc και η Forrester Research Inc. εξετάζουν επίσης τις συναλλαγές και άλλα δομημένα δεδομένα ώστε να υπάρχουν έγκυρα συστατικά στην εφαρμογή της ανάλυσης των Μεγάλων Δεδομένων.

### 3.3 Εργαλεία και τεχνικές ανάλυσης Μεγάλων Δεδομένων

Τα Μεγάλα Δεδομένα μπορούν να αναλυθούν με τα εργαλεία λογισμικού που χρησιμοποιούνται συνήθως ως μέρος των προηγμένων κλάδων αναλυτικών μοντέλων όπως των αναλυτικών μοντέλων προβλέψεων, της άντλησης δεδομένων, των αναλυτικών μοντέλων κειμένου και της στατιστική ανάλυσης. Βασικά εργαλεία

λογισμικού επιχειρηματικής ευφυΐας και εργαλεία απεικόνισης δεδομένων μπορούν επίσης να διαδραματίσουν κάποιο ρόλο στη διαδικασία της ανάλυσης.

Από την άλλη πλευρά υπάρχουν ορισμένοι περιορισμοί. Τα ημι-δομημένα και αδόμητα δεδομένα μπορεί να μην ταιριάζουν καλά στις παραδοσιακές αποθήκες δεδομένων που βασίζονται σε σχεσιακές βάσεις δεδομένων. Επιπλέον, οι αποθήκες δεδομένων ενδέχεται να μην είναι σε θέση να αντεπεξέλθουν στις απαιτήσεις που τίθενται από την επεξεργασία των μεγάλων συνόλων δεδομένων. Για παράδειγμα, υπάρχουν δεδομένα που πρέπει να ενημερώνονται συχνά ή ακόμα και συνεχώς, όπως δεδομένα σε πραγματικό χρόνο σχετικά με την απόδοση των εφαρμογών των κινητών ή των αγωγών πετρελαίου και φυσικού αερίου.

Κατά συνέπεια, πολλοί οργανισμοί που αναζητούν τη συλλογή, επεξεργασία και ανάλυση των μεγάλων δεδομένων έχουν στραφεί σε μια νεότερη κατηγορία τεχνολογιών που περιλαμβάνει Hadoop και τα συναφή εργαλεία, όπως YARN, MapReduce, Spark, Hive, Pigas και βάσεις δεδομένων NoSQL. Οι τεχνολογίες αυτές αποτελούν τον πυρήνα ενός πλαισίου λογισμικού ανοικτού κώδικα που υποστηρίζει την επεξεργασία των μεγάλων και διαφορετικών συνόλων δεδομένων σε συγκεντρωμένα συστήματα.

#### Τεχνικές ανάλυσης Μεγάλων Δεδομένων

Μια μεγάλη ποικιλία από τεχνικές έχουν αναπτυχθεί και προσαρμοστεί για να απεικονίσουν, να αναλύσουν, να χειριστούν και να ομαδοποιήσουν τα Μεγάλα Δεδομένα ώστε να διαχειριστεί ο μεγάλος όγκος τους. Οι τεχνικές αυτές, όπως φαίνεται παρακάτω χρησιμοποιούνται για την ανάλυση των Μεγάλων Δεδομένων και βασίζονται κυρίως στη στατιστική επιστήμη και στην επιστήμη των υπολογιστών (μηχανική μάθηση).

**A / B δοκιμές** - Μια τεχνική στην οποία μια ομάδα ελέγχου συγκρίνεται με ομάδες δοκιμής προκειμένου να καθοριστούν ποιες αλλαγές θα βελτιώσουν μία δεδομένη μεταβλητή, π.χ. το ποσοστό ανταπόκρισης στη στρατηγική μάρκετινγκ. Αυτή η τεχνική είναι επίσης γνωστή ως δοκιμή διάσπασης (split testing) ή δοκιμή κουβά (bucket testing). Ένα παράδειγμα εφαρμογής αυτής της τεχνικής είναι ο προσδιορισμός του κειμένου, των σχεδίων, των εικόνων, των χρωμάτων που θα βελτιώσουν τα ποσοστά συναλλαγών σε μια ιστοσελίδα ηλεκτρονικού εμπορίου. Τα Μεγάλα Δεδομένα επιτρέπουν έναν τεράστιο αριθμό δοκιμών που μπορούν να εκτελεστούν και να

αναλυθούν, εξασφαλίζοντας ομάδες επαρκείς σε μέγεθος που δίνουν τη δυνατότητα εύρεσης στατιστικά σημαντικών διαφορών μεταξύ των ομάδων ελέγχου και δοκιμής. Όταν περισσότερες από μία μεταβλητές εξετάζονται ταυτόχρονα στη συγκεκριμένη δοκιμή, η πολυμεταβλητή γενίκευση αυτής της τεχνικής, η οποία εφαρμόζεται με τη χρήση στατιστικής μοντελοποίησης, συνήθως ονομάζεται «A / B / N» δοκιμή.

**Κανόνες συσχέτισης (Association rule learning)** – Είναι μια σειρά από τεχνικές για την ανακάλυψη σχέσεων μεταξύ των μεταβλητών σε μεγάλες βάσεις δεδομένων, που να έχουν κάποιο νόημα. Οι τεχνικές αυτές αποτελούνται από μια ποικιλία αλγορίθμων για τη δημιουργία και τη δοκιμή πιθανών κανόνων. Μία εφαρμογή αυτής της τεχνικής είναι η ανάλυση καλαθιού, στην οποία ένας λιανέμπορος μπορεί να διαπιστώσει ποια προϊόντα συχνά αγοράζονται μαζί και να χρησιμοποιήσει αυτή την πληροφορία για το μάρκετινγκ. Ένα παράδειγμα που αναφέρεται συχνά είναι η ανακάλυψη ότι πολλοί αγοραστές σούπερ μάρκετ που αγοράζουν πάνες τείνουν επίσης να αγοράζουν και μπίρα. Η συγκεκριμένη τεχνική χρησιμοποιείται για την άντληση δεδομένων.

**Ταξινόμηση** – Είναι μια σειρά από τεχνικές που χρησιμοποιούνται για να προσδιορίσουν σε ποιες κατηγορίες ανήκουν νέα δεδομένα, με βάση ένα σύνολο κανόνων που διέπει την ταξινόμηση των ήδη υπάρχοντων δεδομένων. Μία εφαρμογή αυτής της τεχνικής είναι η πρόβλεψη του ειδικών συμπεριφορών των πελατών (π.χ. αποφάσεις αγοράς, ποσοστό αποσυνδέσεων, ρυθμός κατανάλωσης), όπου υπάρχει μια σαφής υπόθεση ή αντικειμενικό αποτέλεσμα. Επιπλέον, και αυτή η τεχνική χρησιμοποιείται για την άντληση δεδομένων.

**Μέθοδος ανάλυσης συστάδων (Cluster Analysis)** – Είναι μια στατιστική μέθοδος για την ταξινόμηση των αντικειμένων μίας ποικιλόμορφης ομάδα σε μικρότερες ομάδες παρόμοιων αντικειμένων, των οποίων τα χαρακτηριστικά ομοιότητας τους δεν είναι γνωστά εκ των προτέρων. Ένα παράδειγμα της μεθόδου ανάλυσης συστάδων είναι ο διαχωρισμός των καταναλωτών σε παρόμοιες ομάδες για στοχευόμενο μάρκετινγκ. Και αυτή η τεχνική χρησιμοποιείται για την άντληση δεδομένων.

**Άντληση από το πλήθος (Crowdsourcing)** - Μια τεχνική για τη συλλογή στοιχείων που υποβλήθηκαν από μια μεγάλη ομάδα ανθρώπων ή κοινότητας (δηλαδή, το «πλήθος»), μέσω ανοικτής πρόσκλησης, συνήθως μέσω των μέσων κοινωνικής δικτύωσης. Αυτό είναι ένα είδος μαζικής συνεργασίας.



**Ενοποίηση δεδομένων** – Είναι μια σειρά από τεχνικές που χρησιμοποιούνται για την ενοποίηση και ανάλυση δεδομένων από πολλαπλές πηγές, προκειμένου να αναπτυχθούν ιδέες με τρόπους που είναι πιο αποτελεσματικοί και δυνητικά πιο ακριβείς από ότι αν είχαν αναπτυχθεί από την ανάλυση μίας και μόνο πηγής δεδομένων. Για την εφαρμογή ορισμένων τύπων ενοποίησης δεδομένων μπορούν να χρησιμοποιηθούν οι τεχνικές επεξεργασίας σήματος. Ένα παράδειγμα εφαρμογής της συγκεκριμένης τεχνικής είναι τα δεδομένα των αισθητήρων από το Διαδίκτυο των Πραγμάτων που συνδυάζονται για να αναπτυχθεί μια ολοκληρωμένη άποψη για την απόδοση ενός σύνθετου συστήματος, όπως για ένα διυλιστήριο πετρελαίου. Ένα άλλο παράδειγμα, αποτελούν τα δεδομένα από τα μέσα κοινωνικής δικτύωσης, που αναλύονται από την επεξεργασία φυσικής γλώσσας (natural language processing) και μπορούν να συνδυαστούν με στοιχεία πωλήσεων σε πραγματικό χρόνο ώστε να καθοριστεί ποιο αποτέλεσμα θα έχει μια εκστρατεία μάρκετινγκ στην ψυχολογία του πελάτη και την καταναλωτική συμπεριφορά.

**Η άντληση δεδομένων** – Είναι μια σειρά από τεχνικές για την εξαγωγή μοτίβων από μεγάλα σύνολα δεδομένων. Συνδυάζουν μεθόδους από τη στατιστική και τη μηχανική μάθηση με τη διαχείριση βάσεων δεδομένων. Αυτές οι τεχνικές περιλαμβάνουν κανόνες συσχέτισης (association rule learning), ανάλυση συστάδων (cluster analysis) ταξινόμηση και παλινδρόμηση. Οι εφαρμογές αυτής της μεθόδου περιλαμβάνουν άντληση δεδομένων πελατών ώστε να καθοριστούν τα κοινά-στόχοι που είναι πιθανότερο να ανταποκριθούν σε μια προσφορά ή άντληση δεδομένων ανθρώπινου δυναμικού για να προσδιοριστούν τα χαρακτηριστικά των πιο επιτυχημένων εργαζομένων ή ανάλυση καλαθιού για την μοντελοποίηση της αγοραστικής συμπεριφοράς των πελατών.

**Μάθηση συνόλου** - Χρήση πολλαπλών μοντέλων πρόβλεψης (το καθένα έχει αναπτυχθεί χρησιμοποιώντας στατιστική ή / και μηχανική μάθηση) για την επίτευξη καλύτερης πρόβλεψης από ό, τι θα μπορούσε να ληφθεί από οποιοδήποτε συστατικό μοντέλο (constituent model).

**Γενετικοί αλγόριθμοι** – Είναι μια τεχνική που χρησιμοποιείται για τη βελτιστοποίηση και είναι εμπνευσμένη από τη διαδικασία της φυσικής εξέλιξης ή επιβίωσης του ισχυρότερου. Σε αυτήν την τεχνική, οι πιθανές λύσεις κωδικοποιούνται ως χρωμοσώματα που μπορούν να συνδυαστούν και να μεταλλαχθούν. Αυτά τα μεμονωμένα χρωμοσώματα έχουν επιλεγεί για την επιβίωση μέσα σε ένα μοντελοποιημένο περιβάλλον που καθορίζει την καταλληλότητα ή την απόδοση του

κάθε ατόμου στον πληθυσμό. Συχνά περιγράφονται ως ένα είδος εξελικτικού αλγορίθμου. Αυτοί οι αλγόριθμοι είναι κατάλληλοι για την επίλυση μη γραμμικών προβλημάτων. Κάποια παραδείγματα είναι η βελτίωση προγραμματισμού θέσεων εργασίας στον μεταποιητικό τομέα και η βελτιστοποίηση της απόδοσης ενός επενδυτικού χαρτοφυλακίου.

**Μηχανική μάθηση** – Είναι ένας τομέας της επιστήμης των υπολογιστών (μέσα σε ένα πεδίο που ιστορικά ονομάζεται «τεχνητή νοημοσύνη»), ασχολείται με το σχεδιασμό και την ανάπτυξη αλγορίθμων που επιτρέπουν στους υπολογιστές να εξελίσσουν συμπεριφορές που βασίζονται σε εμπειρικά δεδομένα. Μια σημαντική εστίαση της έρευνας μηχανικής μάθησης είναι να μάθουν αυτόματα οι υπολογιστές να αναγνωρίζουν πολύπλοκα μοτίβα και να λαμβάνουν έξυπνες αποφάσεις με βάση τα δεδομένα. Η επεξεργασία φυσικής γλώσσας είναι ένα παράδειγμα της μηχανικής μάθησης.

**Επεξεργασία φυσικής γλώσσας (Natural language processing NLP)** – Είναι μια σειρά από τεχνικές, υποειδικότητα του πεδίου της «τεχνητής νοημοσύνης» και γλωσσολογίας, που χρησιμοποιούν αλγόριθμους υπολογιστών για να αναλύσουν την ανθρώπινη (φυσική) γλώσσα. Πολλές τεχνικές NLP είναι τύποι μηχανικής μάθησης. Μία εφαρμογή του NLP χρησιμοποιεί ανάλυση συναισθήματος στα μέσα κοινωνικής δικτύωσης για να αναγνωρίσει τον τρόπο που οι υποψήφιοι πελάτες αντιδρούν σε μια καμπάνια μίας μάρκας.

**Νευρωνικά δίκτυα** – Είναι υπολογιστικά μοντέλα, εμπνευσμένα από τη δομή και λειτουργία των βιολογικών νευρωνικών δικτύων (δηλαδή, τα κύτταρα και τις συνδέσεις εντός του εγκεφάλου), που βρίσκουν μοτίβα σε δεδομένα. Τα νευρωνικά δίκτυα είναι μία καλά προσαρμοσμένη μέθοδος για την εύρεση μη γραμμικών μοτίβων. Μπορούν να χρησιμοποιηθούν για την αναγνώριση προτύπων και τη βελτιστοποίηση. Παραδείγματα εφαρμογών περιλαμβάνουν τον προσδιορισμό υψηλής αξίας πελατών που υπάρχει κίνδυνος να εγκαταλείψουν μια συγκεκριμένη εταιρεία.

**Βελτιστοποίηση (Optimization)** – Είναι ένα χαρτοφυλάκιο αριθμητικών τεχνικών που χρησιμοποιούνται για τον επανασχεδιασμό πολύπλοκων συστημάτων και διαδικασιών ώστε να βελτιωθούν οι επιδόσεις τους, σύμφωνα με μία ή περισσότερες αντικειμενικές μετρήσεις (π.χ. το κόστος, την ταχύτητα, την αξιοπιστία). Παραδείγματα εφαρμογών περιλαμβάνουν τη βελτίωση των επιχειρησιακών διαδικασιών, όπως τον προγραμματισμό, τα δρομολόγια, τη λήψη στρατηγικών αποφάσεων, το στρατηγικό

χαρτοφυλάκιο προϊόντων, την ανάλυση συνδεδεμένων επενδύσεων, την στρατηγική Έρευνας & Ανάπτυξης του χαρτοφυλακίου. Οι γενετικοί αλγόριθμοι είναι ένα παράδειγμα της τεχνικής αυτής.

**Αναγνώριση προτύπων** – Είναι ένα σύνολο τεχνικών μηχανικής μάθησης που αναθέτουν κάποια τιμή εξόδου (ή ετικέτα) σε μια δεδομένη τιμή εισόδου (ή παράδειγμα), σύμφωνα με ένα συγκεκριμένο αλγόριθμο. Οι τεχνικές ταξινόμησης αποτελούν ένα παράδειγμα.

**Μοντέλα προβλέψεων** – Είναι μια σειρά από τεχνικές με τις οποίες ένα μαθηματικό μοντέλο (που έχει δημιουργηθεί ή επιλεγεί) προβλέπει καλύτερα την πιθανότητα ενός αποτελέσματος. Ένα παράδειγμα μιας εφαρμογής αυτής της μεθόδου είναι η διαχείριση των πελατειακών σχέσεων για την εκτίμηση της πιθανότητας ένας πελάτης να αλλάξει πάροχο ή την πιθανότητα να πωλούνται σε έναν πελάτη και άλλα προϊόντα. Η παλινδρόμηση είναι ένα παράδειγμα από τις πολλές τεχνικές των μοντέλων προβλέψεων.

**Παλινδρόμηση** – Είναι ένα σύνολο από στατιστικές τεχνικές που χρησιμοποιούνται για να προσδιορίσουν πώς αλλάζει η αξία των εξαρτημένων μεταβλητών, όταν μία ή περισσότερες ανεξάρτητες μεταβλητές μεταβάλλονται. Συχνά χρησιμοποιείται για προβλέψεις. Παραδείγματα αυτής της μεθόδου περιλαμβάνουν την πρόβλεψη του όγκου των πωλήσεων που βασίζεται σε διάφορες μεταβλητές της αγοράς και της οικονομίας ή τον προσδιορισμό των μετρήσιμων παραμέτρων που έχουν τη μεγαλύτερη επιρροή στην ικανοποίηση του πελάτη. Αυτή η μέθοδος χρησιμοποιείται για την άντληση δεδομένων.

**Ανάλυση συναισθήματος** - Εφαρμογή της επεξεργασίας φυσικής γλώσσας (natural language processing) και άλλων αναλυτικών τεχνικών για τον εντοπισμό και την εξαγωγή υποκειμενικών πληροφοριών από υλικό κειμένου. Βασικές πτυχές των αναλύσεων αυτών περιλαμβάνουν τον προσδιορισμό του χαρακτηριστικού, της ιδιότητας ή του προϊόντος για το οποίο ένα συναίσθημα εκφράζεται, τον προσδιορισμό του τύπου, της «πολικότητας» (δηλαδή, θετικό, αρνητικό, ή ουδέτερο), του βαθμού και της δύναμης του συναισθήματος. Παραδείγματα εφαρμογών αυτής της μεθόδου περιλαμβάνουν εταιρείες που εφαρμόζουν ανάλυση συναισθήματος για να αναλύσουν τα μέσα κοινωνικής δικτύωσης (π.χ. blogs, microblogs, και κοινωνικά δίκτυα) ώστε να προσδιορίσουν πώς διαφορετικές κατηγορίες πελατών και ενδιαφερομένων ανταποκρίνονται στα προϊόντα τους και στις ενέργειές τους.

**Επεξεργασία σήματος** – Είναι μια σειρά από τεχνικές της μηχανικής ηλεκτρολογίας και των εφαρμοσμένων μαθηματικών που αναπτύχθηκαν αρχικά για την ανάλυση διακριτών και συνεχών σημάτων, δηλαδή, αναπαραστάσεις αναλογικών φυσικών ποσοτήτων (ακόμα και αν αναπαριστώνται ψηφιακά), όπως ραδιοφωνικά σήματα, ήχοι και εικόνες. Αυτή η κατηγορία περιλαμβάνει τεχνικές από τη θεωρία ανίχνευσης σημάτων, η οποία ποσοτικοποιεί την ικανότητα της διάκρισης μεταξύ του σήματος και του θορύβου. Εφαρμογές αυτής της μεθόδου περιλαμβάνουν μοντελοποίηση για ανάλυση χρονολογικών σειρών ή εφαρμογή της σύντηξης δεδομένων για τον καθορισμό μιας πιο ακριβούς ανάγνωσης του συνδυασμού στοιχείων που προέρχονται από λιγότερο ακριβείς πηγές δεδομένων (π.χ. αποσπώντας το σήμα από το θόρυβο).

**Χωρική ανάλυση (Spatial analysis)** – Είναι μια σειρά από τεχνικές, μερικές εφαρμόζονται από τη στατιστική, οι οποίες αναλύουν τις τοπικές, γεωμετρικές, ή γεωγραφικές ιδιότητες που κωδικοποιούνται σε ένα σύνολο δεδομένων. Συχνά, τα στοιχεία για τη χωρική ανάλυση προέρχονται από γεωγραφικά συστήματα πληροφοριών (GIS), που περιλαμβάνουν δεδομένα πληροφοριών εντοπισμού θέσης, π.χ. διευθύνσεις ή γεωγραφικό πλάτος / γεωγραφικό μήκος. Παραδείγματα εφαρμογών αυτής της μεθόδου περιλαμβάνουν την ενσωμάτωση των χωρικών δεδομένων σε χωρικές παλινδρομήσεις (π.χ. πώς συσχετίζεται η προθυμία των καταναλωτών να αγοράσουν ένα προϊόν με την τοποθεσία) ή προσομοιώσεις (π.χ. πώς λειτουργεί το δίκτυο εφοδιαστικής αλυσίδας σε σχέση με τους δικτυακούς τόπους σε διάφορες τοποθεσίες).

**Στατιστική** – Είναι η επιστήμη της συλλογής, της οργάνωσης, και της ερμηνείας των δεδομένων, συμπεριλαμβανομένου του σχεδιασμού των ερευνών και των πειραμάτων. Οι στατιστικές τεχνικές χρησιμοποιούνται συχνά για να αξιολογήσουν τη σχέση μεταξύ μεταβλητών που θα μπορούσε να υπάρχει κατά τύχη («μηδενική υπόθεση»), και τις σχέσεις μεταξύ των μεταβλητών που είναι πιθανό να υπάρχουν ως αποτέλεσμα μίας υποκείμενης αιτιώδους σχέσης (δηλαδή, που είναι «στατιστικά σημαντικές»). Οι στατιστικές τεχνικές χρησιμοποιούνται επίσης για να μειώσουν την πιθανότητα σφαλμάτων τύπου I («ψευδώς θετικά») και τα σφάλματα Τύπου II («ψευδώς αρνητικά»). Ένα παράδειγμα στατιστικής είναι μια δοκιμή A / B για να καθοριστούν ποιες ενέργειες μάρκετινγκ θα αυξήσουν περισσότερο τα έσοδα.

**Προσομοίωση** – Η μοντελοποίηση της συμπεριφοράς πολύπλοκων συστημάτων, που χρησιμοποιούνται συχνά για την πρόβλεψη και τον σχεδιασμό σεναρίων. Οι

προσομοιώσεις Monte Carlo, για παράδειγμα, είναι μια κατηγορία αλγορίθμων που βασίζονται σε επαναλαμβανόμενες τυχαίες δειγματοληψίες, δηλαδή, τρέχουν χιλιάδες προσομοιώσεις, η καθεμία με βάση διαφορετικές υποθέσεις. Το αποτέλεσμα είναι ένα ιστόγραμμα που δίνει μια πιθανή κατανομή των αποτελεσμάτων. Μία εφαρμογή αυτής της μεθόδου είναι η εκτίμηση της πιθανότητας επίτευξης των οικονομικών στόχων δεδομένης της αβεβαιότητας σχετικά με την επιτυχία των διαφόρων πρωτοβουλιών.

**Χρονοσειρές** – Σύνολο τεχνικών και από τη στατιστική επιστήμη αλλά και από την επεξεργασία σήματος για την ανάλυση των αλληλουχιών των σημείων δεδομένων, για την αντιπροσώπευση τιμών σε διαδοχικές περιόδους, για την εξαγωγή ουσιαστικών χαρακτηριστικών από τα δεδομένα. Παραδείγματα ανάλυσης χρονοσειρών είναι η ωριαία αξία ενός χρηματιστηριακού δείκτη ή ο αριθμός των ασθενών που έχουν διαγνωστεί με μια συγκεκριμένη κατάσταση κάθε μέρα. Χρονοσειρές πρόβλεψης είναι η χρήση ενός μοντέλου για την πρόβλεψη μελλοντικών τιμών από μια χρονολογική σειρά που βασίζεται σε γνωστές παρελθοντικές αξίες της ίδιας ή άλλης σειράς. Μερικές από αυτές τις τεχνικές είναι, π.χ. η δομική μοντελοποίηση, η εξαγωγή μίας τάσης από χρονοσειρά, εποχιακών και υπολειμματικών συστατικών, τα οποία μπορεί να είναι χρήσιμα για τον εντοπισμό κυκλικών μοτίβων στα δεδομένα. Παραδείγματα και εφαρμογές περιλαμβάνουν στοιχεία πρόβλεψης πωλήσεων ή πρόβλεψης του αριθμού των ατόμων που θα διαγνωσθούν με μια μολυσματική ασθένεια.

**Μη επιβλεπόμενη μάθηση** – Είναι ένα σύνολο τεχνικών μηχανικής μάθησης που βρίσκει κρυμμένες δομές σε μη ταξινομημένα δεδομένα. Η ανάλυση συστάδων (Cluster Analysis) είναι ένα παράδειγμα εκμάθησης χωρίς επίβλεψη (σε αντίθεση με την εποπτευόμενη μάθηση).

**Εποπτευόμενη μάθηση** – Είναι το σύνολο των τεχνικών μηχανικής μάθησης που συνάγουν μια λειτουργία ή σχέση από ένα σύνολο δεδομένων. Παράδειγμα αυτής της μεθόδου αποτελεί η ταξινόμηση. Αυτή η μέθοδος είναι διαφορετική από την εκμάθηση χωρίς επίβλεψη.

**Οπτικοποίηση** - Τεχνικές που χρησιμοποιούνται για τη δημιουργία εικόνων, διαγραμμάτων ή κινούμενων σχεδίων για να επικοινωνηθούν, να κατανοηθούν και να βελτιωθούν τα αποτελέσματα των Μεγάλων Δεδομένων.

### 3.4 Εμπόδια

Πιθανά εμπόδια που μπορεί να δυσκολέψουν τις πρωτοβουλίες ενός οργανισμού σχετικά με την ανάλυση των Μεγάλων Δεδομένων μπορεί να είναι η έλλειψη δεξιοτήτων ανάλυσης εσωτερικά και το υψηλό κόστος πρόσληψης έμπειρων επαγγελματιών. Παράλληλα, η ποσότητα των πληροφοριών και η ποικιλία, μπορούν επίσης να προκαλέσουν πονοκεφάλους σχετικά με τη διαχείριση των δεδομένων, συμπεριλαμβανομένης της ποιότητας των δεδομένων και τα ζητήματα συνέπειας.

Επίσης, όταν γίνεται ανάλυση των δεδομένων, είναι πολύ σημαντικό να δοθεί προσοχή στα εξής:

- Σύγκριση άσχετων συνόλων δεδομένων μεταξύ τους και εξαγωγή σχέσεων ή ομοιοτήτων
- Ανάλυση ελλιπών ή "βρώμικων" συνόλων δεδομένων και λήψη αποφάσεων με βάση την ανάλυση αυτών
- Ανάλυση συνόλων δεδομένων χωρίς να ληφθούν υπόψη άλλα σημεία δεδομένων που μπορεί να είναι ζωτικής σημασίας
- Ομαδοποίηση σημείων δεδομένων και η αντιμετώπιση τους ως μία ομάδα, ενώ δεν θα έπρεπε. Για παράδειγμα, η αντιμετώπιση των μοναδικών επισκέψεων σε μία ιστοσελίδα και των συνολικών επισκέψεων σαν ένα σύνολο διογκώνουν τον πραγματικό αριθμό των επισκεπτών, αλλά υποτιμούν το πραγματικό ποσοστό συναλλαγών.
- Μπορεί να υπάρχει μία υπόθεση και να αναζητούνται μόνο τα δεδομένα που την υποστηρίζουν και να αγνοούνται όλα τα σημεία δεδομένων που την απορρίπτουν
- Εστίαση σε δεδομένα που είναι άσχετα με το πρόβλημα ή απόσπαση της προσοχής από δεδομένα που δεν είναι άμεσα συνδεδεμένα με τον στόχο της ανάλυσης. Στην εποχή των Μεγάλων Δεδομένων, αυτό είναι σίγουρο ότι θα συμβαίνει όλο και περισσότερο.

- Ανάμιξη της αιτίας ενός φαινομένου με τη συσχέτιση. Αν μια δράση προκαλεί μία άλλη, τότε είναι σίγουρο ότι συσχετίζονται. Αλλά αν δύο πράγματα απλά συμβαίνουν μαζί δεν σημαίνει ότι το ένα προκάλεσε το άλλο, ακόμα κι αν φαίνεται αυτό να έχει νόημα.
- Χρήση συνόλων δεδομένων που είναι πάρα πολύ λίγα για να δείξουν μια τάση ή σύγκριση αποτελεσμάτων που δεν διαφέρουν αρκετά για να έχουν στατιστική σημαντικότητα.
- Εξαγωγή συμπερασμάτων για την λάθος πρόθεση βάσει των δραστηριοτήτων που καταγράφονται στα στοιχεία παρά την προτεινόμενη πρόθεση.

### 3.5 Προκλήσεις

Τα Μεγάλα Δεδομένα έχουν φτιαχτεί για να προσφέρουν τεράστια διορατικότητα στις εταιρείες. Αλλά με τα terabytes και τα petabytes δεδομένων που εισρέουν στους οργανισμούς σήμερα, οι παραδοσιακές αρχιτεκτονικές και υποδομές που υπάρχουν για την αξιοποίησή τους, αποτελούν μέχρι τώρα μία πρόκληση. Οι ομάδες IT επιβαρύνονται με τις ολοένα και αυξανόμενες αιτήσεις για δεδομένα, ad hoc αναλύσεις και εκθέσεις. Τα στελέχη που καλούνται να πάρουν αποφάσεις είναι απογοητευμένα, καθώς πρέπει να περιμένουν ώρες ή και μέρες για να πάρουν απαντήσεις στα ερωτήματά τους.

Πώς μπορούμε να παρουσιάσουμε τα Μεγάλα Δεδομένα με τέτοιο τρόπο ώστε οι ηγέτες των επιχειρήσεων να μπορούν εύκολα και γρήγορα να τα κατανοήσουν και να τα χρησιμοποιήσουν; Η άντληση εκατομμυρίων γραμμών δεδομένων δημιουργεί ένα μεγάλο πονοκέφαλο για τους αναλυτές που είναι επιφορτισμένοι με την ταξινόμηση και παρουσίασή τους. Οι οργανισμοί θα πρέπει να αντιμετωπίσουν πολλές προκλήσεις που σχετίζονται με την οπτικοποίηση των Μεγάλων Δεδομένων, μερικές από τις οποίες περιγράφονται παρακάτω:

- Ταχύτητα

Στο σημερινό άκρως ανταγωνιστικό επιχειρηματικό περιβάλλον, οι εταιρείες δεν πρέπει μόνο να βρουν και να αναλύσουν τα σχετικά δεδομένα που χρειάζονται, θα πρέπει να

τα βρουν και γρήγορα. Η οπτικοποίηση βοηθά τους οργανισμούς να διεξάγουν τις αναλύσεις τους και να λαμβάνουν αποφάσεις πολύ πιο γρήγορα. Ωστόσο, η πρόκληση είναι να λαμβάνονται αποφάσεις με υψηλή ταχύτητα από τον τεράστιο όγκο δεδομένων και το υψηλό επίπεδο λεπτομέρειας.

Η πρόκληση μεγαλώνει καθώς αυξάνει ο βαθμός αναλυτικότητας των δεδομένων. Μια πιθανή λύση είναι ο εξοπλισμός (hardware). Μερικές εταιρείες χρησιμοποιούν αυξημένη μνήμη και ισχυρή παράλληλη επεξεργασία ώστε να επεξεργάζονται μεγάλους όγκους δεδομένων πολύ γρήγορα. Μια άλλη μέθοδος είναι η τοποθέτηση των δεδομένων στη μνήμη, αλλά με μια προσέγγιση του Grid Computing, όπου πολλές μηχανές χρησιμοποιούνται για να λύσουν ένα πρόβλημα. Και οι δύο προσεγγίσεις επιτρέπουν στους οργανισμούς να διερευνούν τεράστιους όγκους δεδομένων και να αποκτούν γνώσεις σε επιχειρηματικό επίπεδο σχεδόν σε πραγματικό χρόνο.

- Κατανόηση δεδομένων

Χρειάζεται μεγάλη κατανόηση για να πάρουν τα δεδομένα τη σωστή μορφή, έτσι ώστε να μπορούν να χρησιμοποιηθούν ως μέρος της ανάλυσης. Για παράδειγμα, εάν τα δεδομένα προέρχονται από το περιεχόμενο των μέσων κοινωνικής δικτύωσης, θα πρέπει να υπάρχει γνώση ποιος είναι ο χρήστης κατά μια γενική έννοια – είναι για παράδειγμα ένας πελάτης που χρησιμοποιεί ένα συγκεκριμένο σύνολο προϊόντων – και κατανόηση του τι είναι αυτό που πρέπει να ανακτηθεί από τα δεδομένα.

Μία λύση σε αυτήν την πρόκληση είναι να υπάρχει η κατάλληλη εμπειρία και γνώση του αντικειμένου. Οι άνθρωποι που αναλύουν τα στοιχεία έχουν μια βαθιά κατανόηση από πού προέρχονται τα δεδομένα, σε τι κοινό απευθύνονται και πως αυτό το κοινό θα ερμηνεύσει αυτές τις πληροφορίες.

- Ποιότητα δεδομένων

Ακόμα κι αν μπορούν να βρεθούν και να αναλυθούν τα δεδομένα γρήγορα και να τοποθετηθούν σε κατάλληλο πλαίσιο για το κοινό για το οποίο προορίζονται, η αξία των δεδομένων για λήψη αποφάσεων θα τεθεί σε κίνδυνο εάν τα στοιχεία δεν είναι ακριβή ή έγκαιρα. Αυτή είναι μια πρόκληση όλων των αναλύσεων δεδομένων, αλλά κατά την εξέταση του όγκου των πληροφοριών που αφορούν τα έργα Μεγάλων Δεδομένων, είναι ακόμα πιο έντονη. Τα αποτελέσματα θα είναι σημαντικά, εάν η ποιότητα των δεδομένων είναι διασφαλισμένη. Για την αντιμετώπιση αυτού του ζητήματος, οι



εταιρείες πρέπει να έχουν μια διακυβέρνηση δεδομένων ή διαδικασία διαχείρισης πληροφοριών που να μπορεί να διασφαλίσει ότι τα δεδομένα θα είναι καθαρά. Είναι πάντα καλύτερο να υπάρχει μια προληπτική μέθοδος για την αντιμετώπιση των ζητημάτων ποιότητας των δεδομένων, ώστε να μην δημιουργηθούν προβλήματα αργότερα.

- Εμφάνιση ουσιαστικών αποτελεσμάτων

Η χάραξη σημείων σε ένα γράφημα για ανάλυση είναι δύσκολο να γίνει όταν πρόκειται να αναλυθούν εξαιρετικά μεγάλες ποσότητες πληροφοριών ή διάφορες κατηγορίες πληροφοριών. Για παράδειγμα, μπορεί να υπάρχουν 10 δισεκατομμύρια γραμμές δεδομένων από κωδικούς λιανικής που πρέπει να συγκριθούν. Είναι πάρα πολύ δύσκολο για τον χρήστη να δει 10 δισεκατομμύρια πεδία στην οθόνη. Ένας τρόπος για να επιλυθεί αυτό το θέμα είναι να ομαδοποιηθούν τα στοιχεία σε υψηλότερο επίπεδο, για παράδειγμα σε επίπεδο μάρκας ή εταιρείας. Έτσι, μικρότερες ομάδες δεδομένων θα είναι ορατές και αναγνώσιμες. Με την ομαδοποίηση είναι ευκολότερο να απεικονιστούν τα δεδομένα πιο αποτελεσματικά.

- Αντιμετώπιση των ακραίων τιμών

Με τις γραφικές παραστάσεις έγινε δυνατή η απεικόνιση των δεδομένων, από την οποία φαίνονται τάσεις και ακραίες τιμές πολύ πιο γρήγορα από ό, τι από τους πίνακες που περιέχουν αριθμούς και κείμενο. Οι χρήστες μπορούν εύκολα να εντοπίσουν τα θέματα που χρειάζονται προσοχή απλά διαβάζοντας ένα γράφημα. Οι ακραίες τιμές αποτελούν συνήθως περίπου το 1-5 τοις εκατό των δεδομένων, αλλά όταν γίνεται εργασία με τεράστιες ποσότητες δεδομένων, η απεικόνιση του 1-5 τοις εκατό των δεδομένων είναι μάλλον δύσκολη. Πώς γίνεται να αντιπροσωπευτούν αυτά τα σημεία χωρίς να υπάρχει εμπλοκή σε θέματα σχεδίασης; Πιθανές λύσεις είναι να αφαιρεθούν οι ακραίες τιμές από τα δεδομένα (και επομένως από το διάγραμμα) ή να δημιουργηθεί ένα ξεχωριστό γράφημα για τις ακραίες τιμές. Μπορούν επίσης να συνδυαστούν τα αποτελέσματα των ακραίων τιμών με την κατανομή των δεδομένων. Ενώ οι ακραίες τιμές δεν μπορούν να είναι αντιπροσωπευτικές των δεδομένων, μπορούν ωστόσο κάποιες φορές να αποκαλύψουν αθέατες και πιθανώς πολύτιμες πληροφορίες.

- Ολοκλήρωση Δεδομένων

Είναι η ικανότητα συνδυασμού των δεδομένων που δεν έχουν παρόμοια δομή ή πηγή, με ταχύτητα και με λογικό κόστος. Με τέτοια ποικιλία, μία σχετική πρόκληση είναι πώς θα διαχειριστεί και πώς θα ελεγχθεί η ποιότητα των δεδομένων, έτσι ώστε να συνδεθούν δεδομένα που είναι κατανοητά με δεδομένα που δεν είναι κατανοητά.

- Όγκος Δεδομένων

Είναι η ικανότητα της διαχείρισης του όγκου με αποδεκτή ταχύτητα, έτσι ώστε οι πληροφορίες να είναι διαθέσιμες στους φορείς λήψης αποφάσεων, όταν το χρειάζονται.

- Διαθεσιμότητα Δεξιοτήτων

Τα Μεγάλα Δεδομένα αξιοποιούνται με νέα εργαλεία και εξετάζονται με διαφορετικούς τρόπους. Υπάρχει έλλειψη ανθρώπων με δεξιότητες για να συγκεντρώσουν στοιχεία, να τα αναλύσουν και να δημοσιεύσουν τα αποτελέσματα ή συμπεράσματα.

- Κόστος Λύσης

Από τη στιγμή που τα Μεγάλα Δεδομένα έχουν ανοίξει έναν κόσμο πιθανών βελτιώσεων στις επιχειρήσεις, υπάρχει μεγάλος πειραματισμός και ανεύρεση νέων σχεδίων και ιδεών που φέρνουν αξία. Για να εξασφαλιστεί η θετική απόδοση της επένδυσης για ένα έργο με Μεγάλα Δεδομένα, είναι σημαντικό να μειωθεί το κόστος των λύσεων που χρησιμοποιούνται για να βρεθεί αυτή η αξία.

### 3.6 Νέες Τάσεις στην Ανάλυση των Μεγάλων Δεδομένων

- Ανάλυση Μεγάλων Δεδομένων στο Σύννεφο (Cloud)

Μία βάση δεδομένων με διαφορετική αρχιτεκτονική από τις σχεσιακές βάσεις δεδομένων είναι η Hadoop. Η Hadoop είναι ένα πλαίσιο και ένα σύνολο εργαλείων για την επεξεργασία πολύ μεγάλων συνόλων δεδομένων. Αρχικά, σχεδιάστηκε για να δουλεύει με ομάδες φυσικών μηχανών, αλλά αυτό έχει αλλάξει καθώς ένας αυξανόμενος αριθμός τεχνολογιών είναι διαθέσιμος για την επεξεργασία των

δεδομένων στο σύννεφο (cloud). Παραδείγματα περιλαμβάνουν το Redshift της Amazon που φιλοξενεί την BI αποθήκη δεδομένων, το BigQuery, υπηρεσία ανάλυσης δεδομένων της Google, η Bluemix πλατφόρμα cloud της IBM, και το Kinesis, υπηρεσία επεξεργασίας δεδομένων της Amazon.

Η Smarter Remarketer, ένας πάροχος αναλυτικών μοντέλων που βασίζονται σε SaaS, τμηματοποίησης και υπηρεσιών μάρκετινγκ, άλλαξε πρόσφατα από μια δομή βάσης δεδομένων in-house, Hadoop και MongoDB, σε Amazon Redshift, μια αποθήκη δεδομένων που βασίζεται σε σύννεφο (cloud). Μία εταιρεία που έχει τη βάση της στην Ινδιανάπολη συλλέγει πωλήσεις που γίνονται τόσο μέσω του διαδικτύου, όσο και μέσω των συμβατικών καταστημάτων, δημογραφικά δεδομένα των πελατών, καθώς και δεδομένα για τη συμπεριφορά σε πραγματικό χρόνο και στη συνέχεια αναλύει τις πληροφορίες αυτές για να βοηθήσει τις εταιρείες λιανικής να δημιουργήσουν στοχευμένα μηνύματα για να κερδίσουν την επιθυμητή ανταπόκριση από την πλευρά των αγοραστών, σε ορισμένες περιπτώσεις, σε πραγματικό χρόνο. Η αλλαγή σε Redshift ήταν πιο αποδοτική για τις ανάγκες δεδομένων της Smart Remarketer, αφού είχε εκτεταμένες αναφορές για δομημένα δεδομένα. Ως εφαρμογή φιλοξενίας, είναι επεκτάσιμη και σχετικά εύκολη στη χρήση.

- Hadoop: Η νέα επιχείρηση του λειτουργικού συστήματος των δεδομένων

Κατανεμημένα αναλυτικά πλαίσια, όπως το MapReduce, εξελίσσονται σε διαχειριστές πόρων που σταδιακά μετατρέπονται σε Hadoop, σε ένα γενικού σκοπού λειτουργικό σύστημα δεδομένων. Τι σημαίνει όμως κάτι τέτοιο για μια επιχείρηση; Όπως η SQL, το MapReduce, η επεξεργασία ροής δεδομένων, τα γραφήματα αναλύσεων και άλλες εργασίες μπορούν να τρέξουν σε Hadoop με επαρκείς επιδόσεις. Περισσότερες επιχειρήσεις θα χρησιμοποιούν το Hadoop ως κόμβο δεδομένων. Το Hadoop θα είναι ένα χαμηλού κόστους, και γενικής χρήσης μέρος για να τοποθετηθούν τα στοιχεία που πρέπει να αναλυθούν, καθώς έχει τη δυνατότητα να τρέξει πολλά διαφορετικά είδη αναζητήσεων και λειτουργιών δεδομένων.

- Περισσότερα προγνωστικά αναλυτικά μοντέλα

Με τα Μεγάλα Δεδομένα, οι αναλυτές έχουν όχι μόνο περισσότερα δεδομένα για να δουλέψουν, αλλά και την επεξεργαστική ισχύ για να χειριστούν τον μεγάλο αριθμό αρχείων με τα πολλά χαρακτηριστικά. Η παραδοσιακή μηχανική μάθηση χρησιμοποιεί στατιστική ανάλυση που βασίζεται σε ένα δείγμα του συνόλου δεδομένων. Τώρα

υπάρχει η δυνατότητα να δημιουργηθεί ένας πολύ μεγάλος αριθμός εγγραφών και πολύ μεγάλος αριθμός χαρακτηριστικών ανά εγγραφή, γεγονός το οποίο αυξάνει την προβλεψιμότητα. Ο συνδυασμός των μεγάλων δεδομένων και της υπολογιστικής ισχύος δίνει τη δυνατότητα στους ερευνητές να διερευνήσουν νέα δεδομένα συμπεριφοράς κατά τη διάρκεια της ημέρας.

Από την άλλη πλευρά, με το Hadoop υπάρχει ένα πρόβλημα που είναι η ταχύτητα. Με Hadoop μπορούν να απαντηθούν τα ερωτήματα έως και 20 φορές πιο αργά από ό, τι με τις πιο καθιερωμένες τεχνολογίες. Έτσι, χρησιμοποιείται σε συνδυασμό με το Apache Spark, μια μεγάλης κλίμακας μηχανή επεξεργασίας δεδομένων, και το συνδεδεμένο εργαλείο αναζήτησης SQL, το Spark SQL. Το Spark προσφέρει γρήγορη διαδραστική αναζήτηση, υπηρεσίες γραφημάτων και δυνατότητες συνεχούς ροής δεδομένων. Τα δεδομένα υπάρχουν στο Hadoop, ωστόσο η επεξεργασία τους γίνεται με Spark, γεγονός που επιτρέπει να γίνεται η επεξεργασία έως και 100 φορές γρηγορότερα.

- SQL σε Hadoop: Ταχύτερη, καλύτερη

Ένας έξυπνος προγραμματιστής και μαθηματικός μπορεί να χρησιμοποιήσει τα δεδομένα που είναι σε Hadoop και να κάνει μία ανάλυση. Ωστόσο, οι περισσότεροι άνθρωποι θα ήθελαν μία γλώσσα προγραμματισμού με την οποία να είναι εξοικειωμένοι και αυτή είναι η SQL. Εργαλεία που υποστηρίζονται με SQL-όπως η αναζήτηση, επιτρέπουν στους επιχειρηματικούς χρήστες, που γνωρίζουν ήδη SQL, να εφαρμόζουν παρόμοιες τεχνικές στα εν λόγω δεδομένα. Η SQL σε Hadoop ανοίγει την πόρτα για να μπει το Hadoop σε μια επιχείρηση, καθώς οι επιχειρήσεις δεν χρειάζεται να κάνουν μια επένδυση σε υψηλά εξειδικευμένους επιστήμονες δεδομένων και αναλυτές.

- Περισσότερες, καλύτερες NoSQL

Εναλλακτικές των παραδοσιακών SQL σχεσιακών βάσεων δεδομένων, που ονομάζονται NoSQL (συντομία του “Not Only SQL”), κερδίζουν γρήγορα δημοτικότητα ως εργαλεία για χρήση σε συγκεκριμένα είδη εφαρμογών ανάλυσης. Υπάρχουν 15 με 20 open-source NoSQL βάσεις δεδομένων, καθμία με τη δική της ειδικευση. Για παράδειγμα, ένα προϊόν NoSQL με δυνατότητα δημιουργίας γραφικών παραστάσεων, όπως το ArangoDB, προσφέρει έναν ταχύτερο, πιο άμεσο τρόπο για την ανάλυση του

δικτύου των σχέσεων μεταξύ πελατών ή πωλητών από ό, τι μια σχεσιακή βάση δεδομένων.

- Βαθιά μάθηση (Deep learning)

Η Βαθιά μάθηση (deep learning), είναι ένα σύνολο τεχνικών μηχανικής μάθησης που βασίζεται σε νευρωνικά δίκτυα, εξακολουθεί να εξελίσσεται και παρουσιάζει μεγάλες δυνατότητες για την επίλυση των προβλημάτων των επιχειρήσεων. Επιτρέπει στους υπολογιστές να αναγνωρίζουν αντικείμενα που παρουσιάζουν ενδιαφέρον σε δεδομένα μεγάλου όγκου, αδόμητα και δυαδικά, και να συμπεραίνουν σχέσεις χωρίς να χρειάζονται συγκεκριμένα μοντέλα ή οδηγίες προγραμματισμού. Για παράδειγμα, ένας αλγόριθμος βαθιάς μάθησης που εξέτασε τα δεδομένα από την Wikipedia έμαθε από μόνος του ότι η Καλιφόρνια και το Τέξας, είναι και οι δύο πολιτείες των ΗΠΑ. Δεν χρειάστηκε να διαμορφωθεί για να κατανοήσει την έννοια του κράτους και της χώρας, και αυτό είναι μία μεγάλη διαφορά μεταξύ της παλαιότερης μηχανικής μάθησης και των αναδυόμενων μεθόδων βαθιάς μάθησης.

Τα Μεγάλα Δεδομένα θα κάνουν πράγματα με πολύ διαφορετικό και αδόμητο κείμενο χρησιμοποιώντας προηγμένες τεχνικές όπως η αναλυτική βαθιάς μάθησης για να βοηθήσουν με τρόπους που μόλις τώρα αρχίζουν να κατανοούνται. Για παράδειγμα, θα μπορούσε να χρησιμοποιηθεί βαθιά μάθηση για να αναγνωριστούν πολλά διαφορετικά είδη δεδομένων, όπως τα σχήματα, τα χρώματα και τα αντικείμενα σε ένα βίντεο - ή ακόμη και την παρουσία μιας γάτας μέσα σε εικόνες. Η έννοια της γνωστικής εμπλοκής, των προηγμένων αναλυτικών μοντέλων και τα πράγματα που αυτό συνεπάγεται είναι μια σημαντική μελλοντική τάση.

## SAP HANA

Πέρα από το Hadoop, υπάρχει η SAP HANA που είναι μια πλατφόρμα στη μνήμη των υπολογιστών που έχει αλλάξει εντελώς την σχεσιακή βάση δεδομένων της βιομηχανίας. Συνδυάζει βάση δεδομένων, επεξεργασία αιτήσεων, καθώς και υπηρεσίες ολοκλήρωσης σε μια ενιαία πλατφόρμα. Η ίδια αρχιτεκτονική παρέχει επίσης βιβλιοθήκες για πρόβλεψη, σχεδιασμό, επεξεργασία κειμένου και αναλυτικά μοντέλα. Η τελευταία έκδοσή της, SPS10 προσφέρει βελτιωμένες δυνατότητες για Μεγάλα Δεδομένα (Big Data) και το Διαδίκτυο των Πραγμάτων. Παράλληλα, παρέχει επιπλέον διαθεσιμότητα/χωρητικότητα και αποκατάσταση των καταστροφών, καθώς και νέα εργαλεία για την ανάπτυξη εξειδικευμένων εφαρμογών.

### Μεγάλα Δεδομένα & Διαδίκτυο των πραγμάτων:

Η νέα έκδοση της SAP HANA προσφέρει τη δυνατότητα συγχρονισμού απομακρυσμένων δεδομένων. Υποστηρίζει το συγχρονισμό δεδομένων μεταξύ του SAP HANA μίας επιχείρησης με χιλιάδες βάσεις δεδομένων στην άκρη του δικτύου, έτσι ώστε να δίνει τη δυνατότητα προληπτικής δράσης και λήψης αποφάσεων με βάση τις πιο πρόσφατες πληροφορίες. Επιπλέον, δίνει τη δυνατότητα συλλογής δεδομένων από το διαδίκτυο των πραγμάτων, από απομακρυσμένες συσκευές, τα οποία μπορούν στη συνέχεια να αναλυθούν σε συνδυασμό με τα στοιχεία των επιχειρήσεων. Το SAP HANA SPS 10 περιλαμβάνει μια πληθώρα νέων δυνατοτήτων για τη ροή δεδομένων όπως είναι το χαρακτηριστικό για «ροή έξυπνων δεδομένων» που επιτρέπει την υποστήριξη τεράστιων όγκων δεδομένων και τη διάδραση με χιλιάδες συσκευές του διαδικτύου πραγμάτων μέσω των πρωτοκόλλων HTTP / REST και WebSocket. Δεύτερον, παρέχει ένα νέο, ελαφρύ επεξεργαστή συνεχούς ροής, το "Streaming Lite", το οποίο μπορεί να αναπτυχθεί σε απομακρυσμένες συσκευές και πύλες του διαδικτύου των πραγμάτων. Αυτός ο επεξεργαστής χρησιμοποιώντας τα έξυπνα στοιχεία ροής μπορεί να κάνει προεργασία γεγονότων και να δημιουργήσει τοπικές ειδοποιήσεις (alerts) με βάση τους κανόνες της επιχείρησης, πριν την αποστολή του μηνύματος στη SAP HANA. Επιπλέον, προσφέρει περισσότερες δυνατότητες για την αποθήκευση δεδομένων με μεγαλύτερη αποτελεσματικότητα. Ακόμη, υποστηρίζει πιο σύνθετες ανάγκες επεξεργασίας.

Σχετικά με το κομμάτι των Μεγάλων Δεδομένων (Big Data), η SAP HANA προσφέρει έξυπνες δυνατότητες ενοποίησης δεδομένων. Για παράδειγμα, προστέθηκαν νέοι προσαρμογείς (adaptors) για αμφίδρομη αντιγραφή δεδομένων και σύνδεση με εξωτερικές πηγές δεδομένων. Για την καλύτερη αξιοποίηση της Hadoop, η SAP HANA SPS 10 δουλεύει με τις πιο πρόσφατες διανομές Hadoop και προσφέρει ταχύτερη, βελτιστοποιημένη μεταφορά δεδομένων προς και από Spark SQL. Οι διαχειριστές μπορούν να διαχειριστούν πλέον HANA και Hadoop συστάδες χρησιμοποιώντας ένα ενιαίο περιβάλλον εργασίας μέσα από SAP HANA και Apache Ambari. Τέλος, η SAP HANA περιλαμβάνει δυνατότητες «έξυπνης ποιότητας» των δεδομένων που καθαρίζουν τα δεδομένα και συγχωνεύουν τις διπλοεγγραφές, πράγμα που επιτρέπει μεγαλύτερη εμπιστοσύνη και συμμόρφωση με όλα τα δεδομένα μίας επιχείρησης.

## 4 Αναλυτικά μοντέλα

Παρακάτω παραθέτονται ορισμένα αναλυτικά μοντέλα που επικεντρώνονται κυρίως στον τομέα των ταχικίνητων καταναλωτικών αγαθών. Τα μοντέλα αυτά είναι το Brand Advantage – Marketing Mix Modelling, η Ανάλυση Καλαθιού, η Βελτιστοποίηση Κωδικολογίου, το Demand Engineered Clustering, οι Συστάδες Καταστημάτων (Store Clustering), τα Store Group Profilers, τα Μοντέλα προβλέψεων (Forecasting), οι Αγοραστικές Περιστάσεις (Shopper Occasions), τα Δίκτυα Προϊόντων (Product Networks) και οι Κανόνες Συσχέτισης (Association Rules).

### 4.1 Brand Advantage - Marketing Mix Modelling

Στη σημερινή εποχή οι βιομηχανίες αντιμετωπίζουν όλο και μεγαλύτερες πιέσεις, τόσο σε τοπικό όσο και σε διεθνές επίπεδο. Καλούνται να μειώσουν τα κόστη ενώ παράλληλα υπάρχουν μεγάλες αλλαγές στα δημογραφικά χαρακτηριστικά των καταναλωτών, εξατομίκευση των προσφορών και κατακερματισμός των μέσων μαζικής ενημέρωσης. Οι λιανέμποροι από την άλλη πλευρά, έρχονται να προσθέσουν άλλα δύο θέματα, την αύξηση των προϊόντων ιδιωτικής ετικέτας και τη συγκέντρωση του λιανεμπορίου.

Για να αντεπεξέλθουν σε αυτό το πολύπλοκο περιβάλλον, οι βιομήχανοι χρειάζονται βαθιά γνώση του μάρκετινγκ και των αναλυτικών εργαλείων. Το Marketing Mix Modelling είναι μία λύση που συνδυάζει ανάλυση, κατανόηση του επιχειρηματικού προβλήματος, δεδομένα και τεχνολογία και βοηθάει τις βιομηχανίες να αυξήσουν την αποδοτικότητα των ενεργειών μάρκετινγκ, των πωλήσεων και της επικοινωνίας. Συνδυάζει προηγμένες αναλυτικές μεθόδους με οικονομικά στοιχεία του πελάτη και μετράει την απόδοση μίας επένδυσης (ROI) όσον αφορά κάθε ενέργεια μάρκετινγκ και συμβάλλει στην βελτιστοποίηση της δαπάνης για κάθε μάρκα και κάθε κωδικό. Η συγκεκριμένη λύση αξιολογεί όλους τους παράγοντες που επηρεάζουν τις πωλήσεις ταυτόχρονα. Ακόμη λαμβάνει υπόψη τον ανταγωνισμό όσον αφορά τη βασική τιμή, την διαφήμιση και τις ενέργειες μέσα στα καταστήματα. Μπορεί επίσης να εντοπίσει την επίδραση που θα έχει στην μάρκα του πελάτη, η διαφήμιση για ένα ανταγωνιστικό προϊόν.

Τα οφέλη αυτής της λύσης για τις βιομηχανίες είναι ο καθορισμός της συνεισφοράς σε όγκο πωλήσεων και της απόδοσης (ROI) κάθε ενέργειας μάρκετινγκ, ο εντοπισμός της αποτελεσματικότητας κάθε ενέργειας ανά χρονική περίοδο και η ποσοτικοποίηση του όγκου που προέρχεται από ενέργειες μάρκετινγκ σε σχέση με τις βασικές πωλήσεις. Δίνει τη δυνατότητα να συνυπολογιστούν τα περιθώρια κέρδους, ώστε να αξιολογηθεί η αποτελεσματικότητα όλης της επιχείρησης, να αξιολογηθούν οι οικονομικές συνέπειες που θα έχουν οι αλλαγές σε κάθε κομμάτι του μάρκετινγκ, καθώς και να προβλεφτούν μελλοντικές πωλήσεις από διαφορετικά πλάνα μάρκετινγκ. Παράλληλα, βοηθάει στην καλύτερη κατανομή των πόρων ανάμεσα στις μάρκες και τους κωδικούς, στην βελτίωση της αποτελεσματικότητας της επένδυσης σε ενέργειες που γίνονται μέσα στα καταστήματα, με δείκτες αποτελεσματικότητας (key performance indices), με ειδοποιήσεις (alerts), με καρτέλες αποτελεσμάτων (scorecards). Τα αναλυτικά μοντέλα μέτρησης της ελαστικότητας τιμής και της αποτελεσματικότητας των προωθητικών ενεργειών συμβάλλουν στην βελτίωση της τιμολογιακής πολιτικής και των προωθητικών ενεργειών, παρέχοντας ακριβή αποτελέσματα κάθε προώθησης και αλλαγής της τιμής. Η συγκεκριμένη υπηρεσία μπορεί να χρησιμοποιηθεί για μοντελοποίηση ολόκληρων αγορών ή συγκεκριμένων μαρκών ή συσκευασιών, κατά περίπτωση ή να ανανεώνεται ανά τακτά χρονικά διαστήματα, ώστε να παρέχει πληροφορίες που να μπορούν να χρησιμοποιηθούν άμεσα.

Τα τυπικά ερωτήματα που απαντώνται από την συγκεκριμένη λύση είναι τα παρακάτω:

- Ο πελάτης εφαρμόζει το βέλτιστο μίγμα μάρκετινγκ;
- Ποιος είναι ο ρόλος της διαφήμισης στο μίγμα μάρκετινγκ;
- Ποια είναι η αποδοτικότητα μίας προσφοράς;
- Ποιος είναι ο ρόλος των προωθητικών ενεργειών μέσα στα καταστήματα;
- Πως μπορεί να προσαρμοστεί το πλάνο μάρκετινγκ ώστε να φέρει περισσότερες πωλήσεις χωρίς επιπλέον κόστος;
- Πόσο αποτελεσματική είναι η δαπάνη για προωθητικές ενέργειες;
- Ποια είναι η απόδοση της επένδυσης (ROI) σε νέα μέσα επικοινωνίας;

#### 4.1.1 Γενικά για το Marketing Mix

Το μοντέλο του Marketing Mix (επίσης γνωστό ως 4 P) μπορεί να χρησιμοποιηθεί από τους μαρκετίερς ως ένα εργαλείο για να βοηθήσει στην εφαρμογή μιας στρατηγικής μάρκετινγκ. Είναι σημαντικό να γίνει κατανοητό ότι οι αρχές του Marketing Mix είναι



ελέγξιμες μεταβλητές. Το μίγμα μάρκετινγκ μπορεί να ρυθμιστεί σε συχνή βάση ώστε να ανταποκρίνεται στις μεταβαλλόμενες ανάγκες της αγοράς-στόχου και τις άλλες δυναμικές του περιβάλλοντος μάρκετινγκ. Τα 4P αποτελούν αναπόσπαστο τμήμα της διαδικασίας στρατηγικού σχεδιασμού του οργανισμού και αποτελούνται από την ανάλυση:

- Προϊόντος (Product)
- Τιμής (Price)
- Τόπου (Place)
- Προώθησης (Promotion)

Η λειτουργία του μίγματος μάρκετινγκ είναι να συμβάλει στην ανάπτυξη ενός πακέτου (mix) που όχι μόνο θα ικανοποιήσει τις ανάγκες των πελατών στις αγορές-στόχους, αλλά ταυτόχρονα θα μεγιστοποιήσει την απόδοση του οργανισμού.

Ποια είναι η κινητήρια δύναμη για την ανάπτυξη του Marketing Mix Modeling (MMM); Για παράδειγμα, η διαφήμιση είναι ένα σημαντικό μέρος των περισσότερων προϋπολογισμών μάρκετινγκ. Δεν είναι έκπληξη το γεγονός ότι παίρνει ένα μεγάλο μέρος από τη χρηματοδότηση. Δείχνοντας ένα σαφές και μετρήσιμο ROI για δαπάνες μάρκετινγκ διαμορφώνεται ο κανόνας, όπως θα έπρεπε να είναι. Το μάρκετινγκ ανταγωνίζεται με άλλες λειτουργίες σε κάθε επιχείρηση για περιορισμένους εταιρικούς πόρους και πρέπει να δικαιολογήσει το μερίδιό του σε αυτούς. Για ορισμένους οργανισμούς το Marketing Mix Modeling (MMM) θα μπορούσε να αποτελέσει ένα πολύτιμο εργαλείο για τη διαχείριση των εξόδων στο σύνολο του μίγματος μάρκετινγκ. Τι κάνει το Marketing Mix Modeling (MMM); Το Marketing Mix Modeling (MMM) εφαρμόζει στατιστικές διαδικασίες για τον προσδιορισμό:

- Των παραγόντων που οδηγούν τις πωλήσεις
- Τη σχετική σημασία του καθενός από αυτούς τους παράγοντες
- Την απόδοση των επενδύσεων για διάφορες δραστηριότητες
- Το βέλτιστο μίγμα των δαπανών σε κάθε μία από τις δραστηριότητες

Τι πρόβλημα προσπαθεί να πραγματευθεί το Marketing Mix Modeling (MMM); Οι πωλήσεις των προϊόντων και υπηρεσιών είναι αποτέλεσμα μιας σύνθετης σειράς παραγόντων, έτσι η μέτρηση της αποτελεσματικότητας του μάρκετινγκ πρέπει να αναγνωρίσει και να κατανοήσει τους παράγοντες αυτούς. Παρακάτω είναι μερικοί από τους βασικούς παράγοντες που έχουν αποδειχθεί ότι επηρεάζουν τις πωλήσεις στον τομέα των καταναλωτικών αγαθών:

- Λιανική τιμή
- Χονδρική τιμή
- Συσκευασία
- Προωθητικές ενέργειες καναλιού
- Εποχικότητα
- Οικονομία
- Καιρός
- Ανταγωνιστικές δραστηριότητες
- Διαφήμιση
- Ενέργειες μέσα στο κατάστημα
- Εκδηλώσεις
- Δραστηριότητες άμεσου μάρκετινγκ

Η κατανόηση της αλληλεπίδρασης μεταξύ αυτών των παραγόντων είναι το κλειδί για την αποτελεσματική διαχείριση των επενδύσεων μάρκετινγκ. Η πολυπλοκότητα του προβλήματος αυτού σημαίνει ότι μπορεί να αντιμετωπιστεί μόνο μέσα από δομημένη στατιστική ανάλυση όπως αυτή προβλέπεται από το Marketing Mix Modeling (MMM).

Τι περιλαμβάνει το Marketing Mix Modeling (MMM); Στη συνέχεια, ακολουθεί η επισκόπηση των βασικών σταδίων:

- Δημιουργία ενός μοντέλου που αναλύει και προβλέπει τις ιστορικές πωλήσεις
- Δοκιμή της ικανότητας πρόβλεψης του μοντέλου σε ένα δείγμα
- Επαναπροσδιορισμός των δεδομένων και πρόβλεψη του μέλλοντος
- Σύγκριση πραγματικών και προβλεπόμενων πωλήσεων και προσδιορισμός των επιπρόσθετων εσόδων
- Εφαρμογή οικονομικών στοιχείων για τον προσδιορισμό της απόδοσης της επένδυσης
- Το μοντέλο δεν επηρεάζεται από μεμονωμένους παράγοντες
- Προσομοίωση των επιπτώσεων των διαφόρων σχεδίων μάρκετινγκ
- Ανάπτυξη/ βελτιστοποίηση των σχεδίων μάρκετινγκ

Ποια είναι τα συνήθη δεδομένα που χρησιμοποιούνται ως πρώτη ύλη για το Marketing Mix Modeling (MMM); Βασικό στοιχείο για την επιτυχία του MMM είναι τα διαθέσιμα εσωτερικά και εξωτερικά δεδομένα στο κατάλληλο επίπεδο λεπτομέρειας. Τα ακόλουθα είναι παραδείγματα των τύπων δεδομένων που χρησιμοποιούνται συνήθως:

- μηνιαία / εβδομαδιαία στοιχεία πωλήσεων συμπεριλαμβανομένων:
  - των προωθητικών ενεργειών που γίνονται μέσα στα καταστήματα
  - της αντίστοιχης πληροφορίας για τους ανταγωνιστές
  - δεδομένα κατά προτίμηση, ACNielsen / IRI ή / και δεδομένα POS από τους λιανέμπορους
- μηνιαία / εβδομαδιαία στοιχεία για δαπάνες διαφήμισης σε εθνικά μέσα ενημέρωσης
  - εθνική / τοπική τηλεόραση, τύπος
  - κατά προτίμηση GRP's
  - ποσό επένδυσης
- εκδηλώσεις Marketing
  - καταναλωτικές προσφορές
  - προωθητικές ενέργειες
  - άλλες σχετικές εκδηλώσεις των καταναλωτών
- άλλες πηγές δεδομένων
  - οικονομία
  - δημογραφικά
  - καιρός

Ποια είναι τα οφέλη του Marketing Mix Modeling (MMM); Η εφαρμογή μιας λύσης MMM θα επιτρέψει σε έναν οργανισμό να:

- μετράει σε τακτική βάση την αποτελεσματικότητα της διαφήμισης και των λοιπών εργαλείων μάρκετινγκ
- βελτιστοποιεί τις δαπάνες μάρκετινγκ μέσα από την κατανόηση των παραγόντων που οδηγούν τις πωλήσεις και των επιπτώσεών τους
- δίνει τη δυνατότητα προσομοίωσης των αλλαγών στο σχέδιο μάρκετινγκ

#### 4.1.2 Σχετικά με τα μέσα ενημέρωσης (Media)

Με την αύξηση της σημαντικότητας των μέσων κοινωνικής δικτύωσης και των διαδικτυακών δημοσίων σχέσεων, παρατηρούνται περισσότερες εταιρείες να αλλάζουν μέθοδο χτισίματος προϋπολογισμού, αναφορών και επένδυσης σε μέσα ενημέρωσης (Media) ώστε να αντανakλούν επαρκώς τις ιστοσελίδες τις οποίες το καταναλωτικό κοινό επισκέπτεται μέσω του διαδικτύου. Η τάση είναι προς την επανεξέταση των επενδύσεων σε τρεις βασικούς τύπους μέσων ενημέρωσης (media), όπως

περιγράφονται παρακάτω, τα οποία δίνουν την ευκαιρία επικοινωνίας με τους καταναλωτές. Κανένας από αυτούς τους τύπους δεν είναι καινούριος αλλά αυτό που είναι καινούριο είναι η αυξανόμενη έμφαση που δίδεται στα ιδιόκτητα (owned) και κερδισμένα (earned) μέσα ενημέρωσης.

Οι βασικοί τύποι μέσων ενημέρωσης που εξετάζονται στα μοντέλα Marketing Mix είναι:

- Τα μέσα ενημέρωσης με πληρωμή (paid media): Είναι τα μέσα μαζικής ενημέρωσης που πληρώνονται ή αγοράζονται ώστε να επικοινωνήσει κάποιος κάτι διαδικτυακά στους επισκέπτες μίας ιστοσελίδας, να έχει πρόσβαση σε πιθανούς πελάτες μέσω της αναζήτησης ή μέσω διαφημιστικών δικτύων προβολής. Τα παραδοσιακά μέσα ενημέρωσης που ανήκουν σε αυτή την κατηγορία είναι ο τύπος, η τηλεόραση, οι διαφημιστικές επιστολές, τα οποία παραμένουν σημαντικά υψηλά στις δαπάνες των εταιρειών για διαφήμιση.
- Τα κερδισμένα μέσα ενημέρωσης (earned media): Παραδοσιακά, τα μέσα που ανήκουν σε αυτή την κατηγορία αφορούν τη δημοσιότητα που δημιουργείται μέσω των δημοσίων σχέσεων. Επενδύουν στοχευμένα στους καθοδηγητές γνώμης που αυξάνουν την ευαισθητοποίηση των καταναλωτών σχετικά με ένα εμπορικό σήμα. Περιλαμβάνουν την επικοινωνία από στόμα σε στόμα που μπορεί να προέλθει από τα μέσα κοινωνικής δικτύωσης και αφορά συνομιλίες σε κοινωνικά δίκτυα, blogs και άλλες κοινότητες. Τα μέσα ενημέρωσης αυτής της κατηγορίας περιλαμβάνουν διάφορες μορφές συνομιλιών που πραγματοποιούνται τόσο διαδικτυακά όσο και παραδοσιακά.
- Τα ιδιόκτητα μέσα ενημέρωσης (owned media): Είναι μέσα μαζικής ενημέρωσης που ανήκουν σε ένα εμπορικό σήμα. Περιλαμβάνουν δικτυακούς τόπους της εταιρείας, blogs, εφαρμογές για κινητά ή κοινωνική τους παρουσία στο Facebook, το Linked In ή το Twitter. Επιπλέον, μπορεί να περιλαμβάνουν φυλλάδια ή καταστήματα λιανικής.

## Πίνακας 1: Τα μέσα ενημέρωσης

Πηγή: Forrester Research, Inc

Τύπος Μέσου	Ορισμός	Παραδείγματα	Ρόλος	Οφέλη	Προκλήσεις
Ιδιότητα Μέσα Επικοινωνίας (Owned Media)	Κανάλι που ελέγχει μία μάρκα	Ιστοσελίδα Blog Λογαριασμός στο Twitter	Φτιάχνονται με σκοπό τη μακροχρόνια σχέση με τους πιθανούς πελάτες και τα κερδισμένα μέσα	Είναι ελέγξιμα Έχουν χαμηλό κόστος Είναι μακροχρόνια Είναι έξυπνα Μπορούν να προσεγγίσουν συγκεκριμένες αγορές (niche audiences)	Δεν υπάρχουν εγγυήσεις Οι πελάτες μπορεί να μην εμπιστεύονται την εταιρική επικοινωνία Χρειάζεται χρόνος να κλιμακωθούν
Μέσα επικοινωνίας επί πληρωμή (Paid Media)	Μία μάρκα πληρώνει για την εκμετάλλευσή ενός καναλιού	Χορηγίες Διαφημίσεις στην τηλεόραση	Συμβάλλουν στην τροφοδότηση των ιδιόκτητων και κερδισμένων μέσων	Σε ζήτηση Έχουν Αμεσότητα Είναι ελέγξιμα	Ακατάστατα Μειούμενα ποσοστά απόκρισης Χαμηλή Αξιοπιστία
Κερδισμένα Μέσα Επικοινωνίας (Earned Media)	Οι πελάτες γίνονται το κανάλι	Επικοινωνία από στόμα σε στόμα	Ακούνε και ανταποκρίνονται-τα κερδισμένα μέσα είναι συχνά αποτέλεσμα της καλής οργάνωσης και εκτέλεσης των ιδιόκτητων μέσων και των μέσων επί πληρωμή	Πιο αξιόπιστα Παίζουν βασικό ρόλο στις πωλήσεις Έχουν διαφάνεια	Μη ελέγξιμα Μπορεί να πάρουν αρνητική διάσταση Δύσκολο να μετρηθούν

Το Brand Advantage – Marketing Mix Modelling είναι μια λύση που προσφέρεται από την IRI η οποία έχει ως κύριο στόχο να ποσοτικοποιήσει την ιστορική επίδραση των διαφόρων μοχλών μάρκετινγκ για τις πωλήσεις των προϊόντων που αναλύθηκαν. Το μαθηματικό εργαλείο είναι οικονομετρικό που επιτρέπει την απομόνωση και τη μέτρηση κάθε μοχλού μάρκετινγκ, προκειμένου να φιλτράρει το "θόρυβο" από τα δεδομένα και να διαχωρίσει τις συνέπειες. Συνήθως η ανάλυση γίνεται σε επίπεδο Μάρκας (εξ ου και ο τίτλος), αλλά υπάρχουν και άλλες περιπτώσεις τα επίπεδα μοντελοποίησης να είναι υπο-μάρκα, τύπος κ.λπ.

Τα παραδοτέα του Brand Advantage – Marketing Mix Modelling περιλαμβάνουν:

#### 1.Data Diagnostics

Αυτή η εφαρμογή δίνει τη δυνατότητα στους πελάτες να έχουν πρόσβαση στις πληροφορίες ανά αλυσίδα.

#### 2.Budget Optimiser

Αυτή η εφαρμογή βρίσκει τη βέλτιστη κατανομή πόρων στα στοιχεία του μίγματος μάρκετινγκ.

#### 3.Performance Drivers

Είναι γραφήματα που συμβάλλουν στην αξιολόγηση της αποτελεσματικότητας των προωθητικών ενεργειών και της τιμής με βάση σημαντικούς παράγοντες που επηρεάζουν την εξέλιξη των πωλήσεων.



#### 4.Planner

Αυτή η εφαρμογή δίνει τη δυνατότητα για δημιουργία σεναρίων σχετικά με τιμές και προωθητικές ενέργειες, καθώς και τις δυνατότητες προσομοίωσης και σύγκρισης σεναρίων. Παρακάτω ακολουθεί μία αναλυτική ενότητα για το Planner.

#### 4.1.3 Planner

Το Planner είναι μια μοναδική διαδικτυακή εφαρμογή, σχεδιασμένη για να καλύψει τις ανάγκες των πελατών σε ένα άκρως ανταγωνιστικό επιχειρηματικό περιβάλλον. Είναι ένα βασικό εργαλείο που επιτρέπει την εξερεύνηση ενός τεράστιου αριθμού πιθανών αποτελεσματικών στρατηγικών μάρκετινγκ και τελικά βοηθάει τον πελάτη να διαφοροποιηθεί από τους ανταγωνιστές του. Με λίγα λόγια, η εφαρμογή Planner είναι ένα γρήγορο, φιλικό προς το χρήστη εργαλείο, το οποίο επιτρέπει την εφαρμογή διαφόρων σεναρίων τιμών και προωθητικών ενεργειών, για ένα ή περισσότερα προϊόντα που ανήκουν στην ίδια κατηγορία. Το αποτέλεσμα αντικατοπτρίζει δυνητικά κέρδη ή ζημίες σε όγκο ή αξία ανάλογα με την προσέγγιση του χρήστη. Από επιχειρηματικής άποψης, θα μπορούσε να θεωρηθεί ως οδηγός προς τον προσδιορισμό μελλοντικών δράσεων.

Τα κύρια δεδομένα εισόδου είναι τα δεδομένα του τελευταίου έτους της ανάλυσης και τα αποτελέσματα από Liquid Modeling (οικονομετρικά μοντέλα). Τα δεδομένα περιλαμβάνουν ενέργειες που αφορούν την τιμή και τις προωθήσεις του προηγούμενου έτους για κάθε προϊόν της κατηγορίας που αναλύεται, ενώ τα αποτελέσματα του μοντέλου Liquid Modelling περιλαμβάνουν τις αντιδράσεις των καταναλωτών στις μεταβολές των τιμών και την ανταπόκρισή τους σε κάθε προωθητικό μηχανισμό. Το μοντέλο που διερευνά την αποτελεσματικότητα των προωθητικών ενεργειών δίνει πολυάριθμες επιλογές στους χρήστες:

- Μπορούν να αποφασίσουν τον αριθμό των εβδομάδων που θα τρέξουν μία προωθητική ενέργεια
- Μπορούν να αποφασίσουν το ποσοστό των καταστημάτων στα οποία θα μπορούσε να εφαρμοστεί η κάθε προωθητική ενέργεια
- Μπορούν να αξιολογήσουν τις επιπτώσεις από την αλλαγή των τιμών ενσωματώνοντας τις επιπτώσεις των ψυχολογικών ορίων τιμής και των κενών από τα ανταγωνιστικά προϊόντα.
- Μπορούν να κάνουν τις παραπάνω κινήσεις σε περισσότερα από ένα προϊόντα, με οποιαδήποτε προτιμώμενο συνδυασμό.
- Μπορούν να αξιολογήσουν τα αποτελέσματα στην αξία πωλήσεων ενός πιθανού σχεδίου μάρκετινγκ.

Άλλες λειτουργίες:

Σενάρια τιμών: Ο χρήστης μπορεί να κάνει σενάρια τιμών για όσα προϊόντα ο ίδιος επιθυμεί και να δει το αποτέλεσμα στα αναφερόμενα προϊόντα (συνήθως τα κύρια προϊόντα της ανάλυσης). Στην περίπτωση που εξετάστηκαν ψυχολογικά όρια τιμής και αποστάσεις τιμής σε σχέση με τα ανταγωνιστικά προϊόντα, η συμβολή τους λαμβάνεται υπόψη. Η τιμή του προϊόντος μπορεί να μειωθεί ή να αυξηθεί κατά 50%.

Σενάρια προωθητικών ενεργειών: Ο χρήστης μπορεί να αλλάξει την κατανομή και τον αριθμό των εβδομάδων που πραγματοποιήθηκαν προωθητικές ενέργειες κατά το τελευταίο έτος και είχαν σημαντική συνεισφορά στις πωλήσεις. Ο αυτοματισμός χρησιμοποιεί τα αποτελέσματα Liquid Modelling (LM) και δείχνει το αποτέλεσμα στα αναφερόμενα προϊόντα.

#### 4.2 Ανάλυση καλαθιού

Στην σημερινή εποχή που όλα έχουν να κάνουν με τον αγοραστή, η μεγάλη πρόκληση για τις βιομηχανίες και τους λιανέμπορους είναι να επικεντρωθούν στους διάφορους τύπους αγοραστών, οι οποίοι έχουν διαφορετικές ανάγκες ανά αγοραστικό ταξίδι. Οι αγοραστές αγοράζουν ένα καλάθι, όχι μεμονομένα προϊόντα. Έτσι, οι λιανέμποροι και οι βιομήχανοι που θέλουν να κερδίσουν μερίδια αγοράς θα πρέπει να καταλάβουν και να βελτιστοποιήσουν το καλάθι αυτό για να κερδίσουν και να διατηρήσουν αγοραστές.

Η ανάλυση καλαθιού πραγματεύεται τις ανάγκες του πελάτη, παρέχοντας σε βάθος ανάλυση και διορατικότητα (insights) σχετικά με το καλάθι του αγοραστή και τις σχετικές ανάγκες που μπορεί να εξυπηρετεί, για παράδειγμα το συγκεκριμένο αγοραστικό ταξίδι. Κατανοώντας τα κίνητρα πίσω από το καλάθι, κατανοούνται ποιες κατηγορίες ή ποια προϊόντα αλληλεπιδρούν και πως η αγορά του ενός επηρεάζει την αγορά του άλλου. Ακόμη, υπάρχουν οι πληροφορίες σχετικά με την τιμολόγηση, το κωδικολόγιο των προϊόντων και τις προωθητικές ενέργειες. Παράλληλα, η ανάλυση καλαθιού συμβάλλει στην κατηγοριοποίηση των καταστημάτων βάσει της αγοραστικής συμπεριφοράς και αυτό δίνει τη δυνατότητα για διαφοροποιημένες ενέργειες μάρκετινγκ, καθώς οι ενέργειες που είναι αποτελεσματικές σε ένα κατάστημα, μπορεί να μην είναι το ίδιο αποτελεσματικές σε ένα άλλο.



Η ανάλυση καλαθιού περιλαμβάνει δεδομένα από συναλλαγές και συγκεκριμένα από τις αποδείξεις πληρωμής που δίνουν οι λιανέμποροι στους αγοραστές κάθε φορά. Οι πληροφορίες που δίνονται από την ανάλυση είναι η χρονική στιγμή της αγοράς, το μέγεθος του καλαθιού, οι σχετιζόμενες προωθητικές ενέργειες, οι ενδείξεις που μπορεί να υπάρχουν ότι κάποια προϊόντα αγοράζονται με κάποια άλλα και πότε γίνεται αυτό. Καλύπτουν όλες τις κατηγορίες που μπορεί να πωλούνται από έναν λιανέμπορο, ακόμη και φρέσκα προϊόντα για τα οποία είναι δύσκολη η παρακολούθηση των πωλήσεών τους.

Τα οφέλη που έχει μία επιχείρηση από την ανάλυση καλαθιού είναι ο εντοπισμός των αναγκών του αγοραστή που επιτρέπει καλύτερη στόχευση και προσαρμοσμένη τιμολογιακή πολιτική. Την βελτίωση των πλάνων προωθήσεων βλέποντας ποιες κατηγορίες είναι συμπληρωματικές η μία με την άλλη και ποιες είναι υποκατάστατες. Μία επιχείρηση μπορεί να έχει επίσης το πιο κερδοφόρο αγοραστικό ταξίδι, βελτιστοποιώντας το κωδικολόγιο των προϊόντων μίας κατηγορίας, τις τιμές και τις προσφορές. Μπορεί να κάνει διαφοροποίηση των στρατηγικών ανά ομάδα καταστημάτων ώστε να ικανοποιηθούν οι διαφορετικές αγοραστικές ανάγκες, καθώς και να αποφύγει την κατάργηση προϊόντων που μπορεί να μην είναι κερδοφόρα αλλά να αποτελούν κίνητρα για μεγαλύτερα καλάθια.

Τυπικές ερωτήσεις που μπορούν να απαντηθούν με την συγκεκριμένη λύση είναι ποιος αγοράζει τα προϊόντα, αν υπάρχουν διαφορετικά αγοραστικά κοινά που ανταποκρίνονται διαφορετικά στις μειώσεις τιμών, τι συμβαίνει όταν γίνεται μία προώθηση, έρχονται νέοι πελάτες ή οι υπάρχοντες πελάτες αποθηκεύουν προϊόν. Τι άλλο αγοράζουν μαζί με τα συγκεκριμένα προϊόντα και πως αυτό θα μπορούσε να επηρεάσει τη στρατηγική μάρκετινγκ. Γιατί μειώνονται οι πωλήσεις. Χρειάζεται να διαφοροποιηθούν οι ώρες που κάποια καταστήματα θα είναι ανοιχτά ή το κωδικολόγιο των προϊόντων. Τι θα γίνει αν θα μειωθεί μία κατηγορία, πως θα επηρεαστούν οι άλλες κατηγορίες. Ποιες είναι οι σημαντικές κατηγορίες για τα μεγάλα καλάθια. Αν γίνει προώθηση σε ένα προϊόν, θα επηρεαστούν άλλες κατηγορίες ή άλλα προϊόντα.

#### 4.3 Βελτιστοποίηση κωδικολογίου

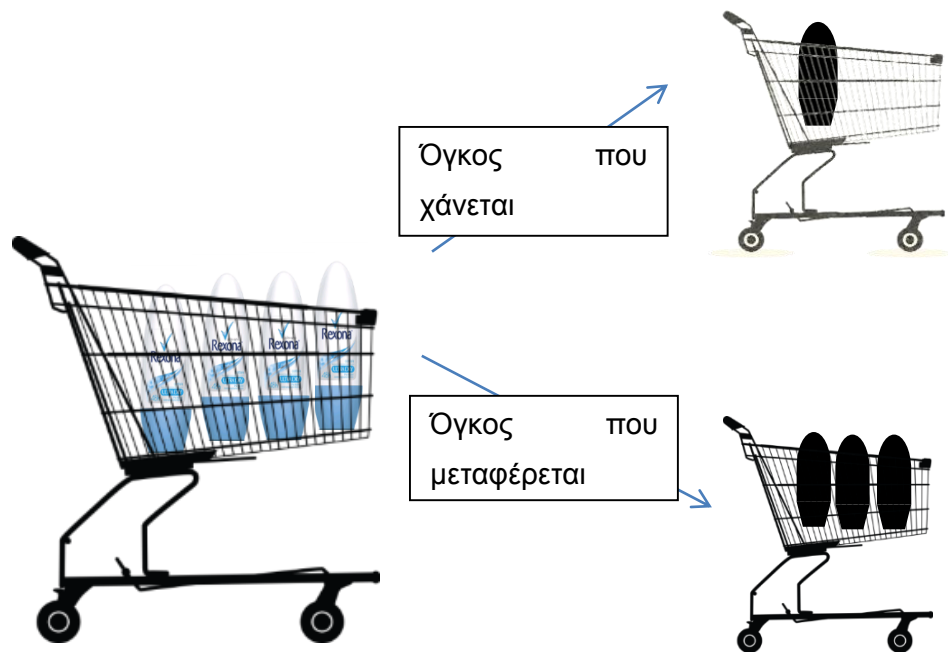
Καθώς όλο και περισσότεροι κωδικοί εισέρχονται στα σουπερ μάρκετ, οι εταιρείες αντιμετωπίζουν ορισμένες προκλήσεις όπως: Έχουν το βέλτιστο κωδικολόγιο για να

πετύχουν τους στόχους τους σε όγκο και αξία; Βοηθάει το κωδικολόγιο τους την βελτιστοποίηση της διαχείρισης της εφοδιαστικής αλυσίδας και κρατάει χαμηλά τα σχετικά κόστη; Πως μπορούν να εντοπιστούν οι καταναλωτικές ανάγκες και να καλυφθούν ώστε να αυξηθεί η κάλυψη στην αγορά; Πότε συγκεκριμένα προϊόντα χάνουν/κερδίζουν ράφι; Πως επηρεάζονται όλα τα διαθέσιμα προϊόντα του εύρους;

Το αναλυτικό μοντέλο βελτιστοποίησης κωδικολογίου παίζει σημαντικό ρόλο στην αύξηση των μεριδίων αγοράς, ιδιαίτερα στη σημερινή εποχή που οι αγοραστές είναι πιο έξυπνοι από ποτέ, βομβαρδίζονται με τεράστιο όγκο πληροφοριών και έχουν ελάχιστο χρόνο να αποφασίσουν μέσα στο κατάστημα ποιο προϊόν θα αγοράσουν. Ένα βέλτιστο κωδικολόγιο καλύπτει τις ανάγκες των καταναλωτών καλύτερα, συμβάλλει στη μείωση του κόστους και στην πιο αποτελεσματική διαχείριση της εφοδιαστικής αλυσίδας. Το συγκεκριμένο μοντέλο αξιολογεί την πραγματική αξία κάθε κωδικού και βοηθάει τις επιχειρήσεις να λάβουν αποφάσεις σχετικά με το βέλτιστο κωδικολόγιο, κάτι που μειώνει τα κόστη και αυξάνει ταυτόχρονα την απόδοση της επένδυσης (ROI). Η απόσυρση ή όχι ενός κωδικού ή ακόμη και η δημιουργία ενός νέου κωδικού με συγκεκριμένα χαρακτηριστικά είναι πλέον μία διαδικασία που βασίζεται σε δεδομένα. Υπάρχουν εργαλεία που επιτρέπουν προσομοίωση σεναρίων με διαφορετικά κωδικολόγια και χώρο στο ράφι με τη χρήση δεδομένων EPOS. Ακόμη, οι συνέπειες προσθήκης ή αφαίρεσης ενός κωδικού ή περισσότερων ή μίας μάρκας ή μίας υποκατηγορίας από μία προϊόντική κατηγορία μπορούν να μετρηθούν πλέον με μεγάλη ακρίβεια.

Πως δουλεύει;

Κάθε προϊόν έχει συγκεκριμένα χαρακτηριστικά όπως μάρκα, υποκατηγορία, συσκευασία, τύπο, όγκο, άρωμα, τιμή. Από αυτά τα χαρακτηριστικά (σε συνδυασμό με τα χαρακτηριστικά του μίγματος μάρκετινγκ) καθορίζεται η ελκυστικότητα του προϊόντος. Η σημαντικότητα του κάθε χαρακτηριστικού καθορίζεται από το μοντέλο. Ο αγοραστής πραγματοποιεί μία αγοραστική απόφαση με βάση την ελκυστικότητα του κάθε χαρακτηριστικού σε σχέση με την ελκυστικότητα των χαρακτηριστικών των άλλων προϊόντων του ίδιου εύρους. Κάθε προϊόν φέρνει έναν επιπλέον όγκο στην κατηγορία, ο οποίος χάνεται στην περίπτωση που αυτό το προϊόν αποσυρθεί. Επιπλέον, το ίδιο προϊόν έχει και έναν μεταβιβάσιμο όγκο, ο οποίος όταν το προϊόν αποσυρθεί, μοιράζεται στα προϊόντα της κατηγορίας. Ο πρώτος επηρεάζει το μέγεθος της αγοράς, ενώ ο δεύτερος επηρεάζει τα μερίδια αγοράς.



Σχήμα 5: Δείχνει τις μονάδες όγκου που χάνονται όταν βγαίνει ένα προϊόν από την κατηγορία και τις μονάδες όγκου που μεταφέρονται στα υπόλοιπα προϊόντα

Τα στάδια αυτής της προσέγγισης είναι τα παρακάτω:

- 1.Επεξεργασία των δεδομένων ανά καταστηματο-εβδομάδα και των πληροφοριών για τα χαρακτηριστικά των προϊόντων σε συνδυασμό με το μίγμα μάρκετινγκ.
- 2.Καθορισμός της σημαντικότητας των χαρακτηριστικών με τη χρήση προηγμένων τεχνικών μοντελοποίησης.
- 3.Χαρτογράφηση σχεδίων και προσπάθεια ερμηνείας των δεδομένων, εντοπίζοντας ομοιότητες και διαφορές.
4. Εκτίμηση του επιπλέον όγκου που το κάθε προϊόν φέρνει στην κατηγορία
- 5.Προσομείωση σεναρίων και καθορισμός των αποτελεσμάτων που έχει το νέο κωδικολόγιο.

Τα οφέλη για τις επιχειρήσεις είναι ότι είναι μία προσέγγιση βασισμένη σε δεδομένα που συμβάλει στην επίτευξη μέγιστων εσόδων, όγκου ή κέρδους μέσω της βελτιστοποίησης του κωδικολογίου. Επιτρέπει την παρακολούθηση των αποτελεσμάτων σε επίπεδο καταστήματος, εβδομάδας και προϊόντος. Δημιουργεί λύσεις όπου κερδίζουν τόσο οι βιομήχανοι όσο και οι λιανέμποροι. Μεγιστοποιεί το αποτέλεσμα που θα έχει η επιχείρηση από την απόσυρση ή προσθήκη κωδικών. Εντοπίζει ανάγκες των καταναλωτών, αυξάνοντας τη γνώση των δυναμικών της κατηγορίας, ενισχύοντας την κάλυψη της αγοράς. Ο εντοπισμός των δυναμικών της κατηγορίας επιτρέπει επίσης την εύρεση υποκατηγοριών και δίνει την προοπτική για μία άλλη ομαδοποίηση στα ράφια.

Από τη συγκεκριμένη λύση, φαίνεται ποια προϊόντα αλληλεπιδρούν από την οπτική γωνία του καταναλωτή μέσω της χαρτογράφησης των προϊόντων. Στην χαρτογράφηση φαίνεται επίσης που υπάρχει κενό για εισαγωγή νέου προϊόντος και ποια προϊόντα είναι περισσότερο ανταγωνιστικά μεταξύ τους. Ακόμη, δείχνει τον κανιβαλισμό που προκαλεί ένα νέο λανσάρισμα στη μάρκα αλλά και την επίδραση που έχει στις υπόλοιπες μάρκες της επιχείρησης αλλά και στο σύνολο της κατηγορίας. Επιπλέον, φαίνεται ποια χαρακτηριστικά είναι σημαντικά για τον καταναλωτή, στην προτίμηση και στην αντίληψή του ώστε να επικοινωνούνται πιο έντονα κατά την προώθηση των προϊόντων.

Παράλληλα, από το μοντέλο της βελτιστοποίησης του κωδικολογίου φαίνεται αν είναι κάποιοι κωδικοί περισσότερο αποδοτικοί σε κάποιες γεωγραφικές περιοχές ή ομάδες καταστημάτων, ποιο είναι το καλύτερο επίπεδο διανομής για κάθε προϊόν ανά λιανέμπορο και τύπο καταστήματος. Επιπλέον, φαίνεται αν αποσυρθεί ένα προϊόν, πως μοιράζεται ο όγκος του στα υπόλοιπα προϊόντα της κατηγορίας και ποια μάρκα θα επηρεαστεί περισσότερο. Από την ποσοτικοποίηση του όγκου που μεταβιβάζεται και του όγκου που δεν μεταβιβάζεται, μπορούν οι λιανέμποροι να βλέπουν ποια προϊόντα τους συμφέρει να βάλουν στα καταστήματά τους, ποια αναπτύσσουν δηλαδή την κατηγορία μέσω του επιπλέον όγκου που φέρνουν (ο οποίος δεν μεταβιβάζεται). Με την προσθήκη εκείνων των προϊόντων που φέρνουν επιπλέον όγκο στην κατηγορία, πετυχαίνεται ταυτόχρονα και μεγιστοποίηση των εσόδων και της κερδοφορίας.

#### 4.4 Demand Engineered Clustering

Τα καταστήματα – σημεία πώλησης ανάλογα με την τοπικότητα και τα καταστήματα του ανταγωνισμού που βρίσκονται στον περίγυρό τους αναπτύσσουν διαφορετικές συμπεριφορές. Έτσι, οι βιομήχανοι και οι λιανέμποροι έρχονται αντιμέτωποι με τις ακόλουθες προκλήσεις: ποια καταστήματα είναι κάτω από τον μέσο όρο της περιοχής στην οποία ανήκουν; Ποια χαρακτηριστικά διαφοροποιούν τις πωλήσεις των καταστημάτων; Πως τα καταστήματα θα μπορούσαν να ομαδοποιηθούν ανάλογα με τα χαρακτηριστικά των πωλήσεών τους;

Η αγοραστική συμπεριφορά διαφέρει από κατάσταση σε κατάσταση, καθώς η ζήτηση για προϊόντα παρουσιάζει διακυμάνσεις ανάλογα με την τοπικότητα. Ορισμένοι παράγοντες που επηρεάζουν τις πωλήσεις ενός καταστήματος είναι ο χρόνος που χρειάζονται οι αγοραστές για να φτάσουν στα σημεία πώλησης, το μέγεθος του καταστήματος, τα κοινωνικά και δημογραφικά χαρακτηριστικά, ο αριθμός των καταστημάτων του ανταγωνισμού που βρίσκονται στη γύρω περιοχή. Έτσι, δημιουργήθηκε η λύση για την ομαδοποίηση των καταστημάτων ανάλογα με συγκεκριμένα χαρακτηριστικά που επιτρέπει προσαρμοσμένες στρατηγικές, οι οποίες στοχεύουν πιο αποτελεσματικά στην ικανοποίηση των αναγκών των αγοραστών. Παράλληλα, βοηθάει τους λιανέμπορους και τους βιομήχανους να μεγιστοποιήσουν την αύξηση των πωλήσεων τους και την κίνηση στα καταστήματα, καθώς και την μείωση του κόστους.

Η λύση Demand Engineering Clustering αφορά την ομαδοποίηση σε επίπεδο καταστήματος όλης της αγοράς που βασίζεται στην ζήτηση του κάθε προϊόντος μίας κατηγορίας. Οι ομάδες των καταστημάτων εντοπίζονται από παρόμοια προφίλ ζήτησης μέσα σε μία κατηγορία. Σε αυτές τις ομάδες μπορούν να εφαρμοστούν στρατηγικές προσαρμοσμένες στον τοπικό χαρακτήρα του κάθε καταστήματος. Για παράδειγμα, η αύξηση ή η μείωση ραφίου και η μείωση των κενών σε σχέση με τον ανταγωνισμό μπορούν να βασιστούν σε αποφάσεις που θα ληφθούν βάσει κάποιων δεδομένων, με μετρήσιμα αποτελέσματα.

Τα οφέλη της συγκεκριμένης λύσης είναι η άντληση πληροφοριών για την περιοχή με τα ίδια χαρακτηριστικά και δημιουργία προφίλ περιοχής. Κατανόηση των χαρακτηριστικών που οδηγούν την ανάπτυξη στην συγκεκριμένη περιοχή. Εντοπισμός

μίας στρατηγικής ανάπτυξης που θα εκμεταλλεύεται τις διαφορές που υπάρχουν στην ζήτηση ανάλογα με την περιοχή. Δημιουργία μίας διαδικασίας, η οποία θα επιβεβαιώνει την ομοιότητα της αγοραστικής συμπεριφοράς και θα μπορεί να εφαρμοστεί αποτελεσματικά ανά ομάδα καταστημάτων/περιοχή. Εντοπισμός των καταστημάτων όπου έχει μεγαλύτερο νόημα να γίνονται προσαρμογές τιμών και προωθητικές ενέργειες. Εντοπισμός των καταστημάτων που αναπτύσσονται με μικρότερο ρυθμό από τον μέσο όρο της περιοχής και σε σχέση με τον ανταγωνισμό, ανακάλυψη των αιτιών και διορθωτικές ενέργειες. Αξιολόγηση αν μία μάρκα έχει τον χώρο ραφίου που της αναλογεί. Διερεύνηση αν υπάρχει κάποια τάση που δεν έχει εντοπιστεί ακόμη.

Που χρησιμεύει

Στο ανταγωνιστικό και δυναμικό περιβάλλον που επικρατεί σήμερα στο λιανεμπόριο, οι λιανέμποροι πρέπει να διακρίνουν τους εαυτούς τους για να αποκτήσουν την αιχμή που είναι απαραίτητη για να κερδίσουν τους κατάλληλους πελάτες και να αυξήσουν το μερίδιο αγοράς τους. Οι περισσότεροι έμποροι λιανικής πώλησης έχουν ένα εμπορικό σήμα, μία στρατηγική μάρκετινγκ και μία ποικιλία. Η πολυπλοκότητα, που δημιουργεί θέμα, είναι ότι τα αποθέματα, οι ρυθμίσεις αποθήκευσης και το σχέδιο προώθησης στα καταστήματα λιανικής πώλησης γίνονται με βάση το ποσοστό πωλήσεων και όχι με βάση τον διαφορετικό τρόπο ζωής των καταναλωτών και τα χαρακτηριστικά της αγοράς. Μια γενική στρατηγική δεν ταιριάζει σε μεμονωμένα καταστήματα με διαφορετικές αγοραστικές συνήθειες καταναλωτών, κατά συνέπεια μια ενιαία λιανική προσφορά γίνεται ένα ανταγωνιστικό μειονέκτημα. Ορισμένες από τις αποφάσεις, όπως η εμπορευματοποίηση και η ποικιλία λαμβάνονται σε επίπεδο καταστήματος, αλλά η διαφοροποίηση μαρκών, προϊόντων και προωθητικών ενεργειών για ορισμένα καταστήματα είναι κάτι ακριβό ή δύσκολο για πολλούς λιανέμπορους. Ως εκ τούτου, προκύπτει η ανάγκη για την κατηγοριοποίηση των καταστημάτων με παρόμοια χαρακτηριστικά σε ομάδες, έτσι ώστε τα καταστήματα με παρόμοια χαρακτηριστικά να στοχεύονται με συγκεκριμένο τρόπο.

#### 4.5 Συστάδες Καταστημάτων (Store Clustering)

Η ανάλυση συστάδων καταστημάτων είναι μια δημοφιλής αναλυτική τεχνική που εξυπηρετεί την κατηγοριοποίηση των καταστημάτων. Είναι μια διαδικασία ομαδοποίησης που ταξινομεί τα «όμοια» καταστήματα σε ένα ενιαίο σύμπλεγμα, που

είναι ταυτόχρονα «ανόμοια» προς τα καταστήματα που ανήκουν σε άλλες ομάδες. Χρησιμοποιείται για την δημιουργία κατάλληλων ομάδων καταστημάτων, τα οποία θα είναι ομοιογενή σε ορισμένα χαρακτηριστικά όπως παρόμοια απόδοση, τμήματα αγοραστών, χαρακτηριστικά λειτουργίας, μέγεθος/τύπος καταστήματος και δημογραφικά χαρακτηριστικά, έτσι ώστε παρόμοια κοινά αγοραστών να στοχεύονται με το ίδιο σχέδιο μάρκετινγκ.

Πως δουλεύει

Η οργάνωση των καταστημάτων σε συστάδες, λαμβάνει υπόψη διάφορους παράγοντες για την ομαδοποίηση. Υπάρχουν δύο τύποι συστάδων καταστημάτων: με βάση τις επιδόσεις και το αντίθετο.

1. Ομαδοποίηση με βάση τις επιδόσεις : Καταστήματα με παρόμοιες επιδόσεις πωλήσεων ομαδοποιούνται. Για παράδειγμα, ένα μικρό κατάστημα με υψηλές πωλήσεις εμπορεύεται με διαφορετικό τρόπο από ένα μικρό κατάστημα με χαμηλές πωλήσεις, οπότε αυτά δύο καταστήματα ανήκουν σε δύο διαφορετικές ομάδες.

2. Ομαδοποίηση που δεν βασίζεται στις επιδόσεις: Τα καταστήματα ομαδοποιούνται ανάλογα με χαρακτηριστικά, όπως:

- Το κλίμα
- Το μέγεθος καταστήματος
- Τον τύπος καταστήματος
- Τους τύπους τοποθεσιών (σε Mall, ανεξάρτητο κατάστημα)

Δημογραφικά στοιχεία των πελατών, όπως:

- Εθνικότητα
- Ηλικιακή ομάδα
- Εισόδημα
- Τρόπος ζωής

Οι έμποροι λιανικής πώλησης για να κατανοήσουν και να εφαρμόσουν Συστάδες Καταστημάτων, πρέπει κάθε συστάδα να έχει ελεγχτεί με βάση τα εξής:

1. Προφίλ των αγοραστών στις συστάδες από τα δεδομένα:

- Αγοραστική συμπεριφορά πελατών
- Μοτίβα δαπανών

- Χαρακτηριστικά του τρόπου ζωής
- Προτιμώμενα προϊόντα
- Προτιμώμενες μάρκες
- Μοτίβα για περιστασιακά ψώνια (δηλαδή πότε ένας πελάτης αγοράζει περισσότερο, για παράδειγμα κατά τη διάρκεια ενός φεστιβάλ, μίας γιορτής;)

## 2. Αναγνώριση σημαντικών αγοραστών μίας κατηγορίας ή μίας μάρκας.

Τα δεδομένα σχετικά με την κατηγορία των πωλήσεων εντός συγκεκριμένων καταστημάτων είναι το κλειδί για πληροφορίες. Ωστόσο, η πληροφορία αυτή δεν αποκαλύπτει ορισμένους αγοραστές σε κάθε ομάδα. Υπάρχουν λίγοι καταναλωτές οι οποίοι έχουν τη μεγαλύτερη πιθανότητα να αγοράσουν προϊόντα που οδηγούν σε πρωτόγνωρες πωλήσεις. Αυτές οι αθέατες ευκαιρίες πωλήσεων στο εσωτερικό κάθε ομάδας μπορούν να εντοπιστούν με τη συλλογή στοιχείων νοικοκυριών, αν υπάρχουν. Αυτά τα δεδομένα χρησιμοποιούνται για να σχεδιαστεί ένα δημογραφικό προφίλ και ο τρόπος ζωής των βαρέων χρηστών των εν λόγω προϊόντων.

## 3. Προσδιορισμός της προοπτικής για κάθε τύπο προϊόντος και μάρκας στην κατηγορία σε κάθε ομάδα.

- Ανάλυση πώς κάθε προϊόν, καθώς και μάρκα συμπεριφέρονται μέσα στην κατηγορία
- Ανάλυση πόσο κάθε τύπος προϊόντος και μάρκας συμβάλλουν στα συνολικά έσοδα

### 4. i. Αναγνώριση των χαρακτηριστικών των μέσων ενημέρωσης των πιστών πελατών.

Ο τρόπος μέσω του οποίου ο πελάτης γνωρίζει το προϊόν ή τα μέσα διαφήμισης στα οποία δίνει προσοχή (εφημερίδες, περιοδικά, διαφημίσεις στην τηλεόραση, το ραδιόφωνο, πινακίδες (banners), διαφημίσεις που αποστέλλονται μέσω των κινητών τηλεφώνων και φέιγ βολάν).

### ii. Διατύπωση στρατηγικής προώθησης για τους πιστούς πελάτες

- Προώθηση προϊόντος / μάρκας μέσω των κατάλληλων καναλιών
  - Στοχοποίηση των σωστών πελατών με τις σωστές προωθητικές ενέργειες (διαμόρφωση στρατηγικών προώθησης για την τόνωση του ενδιαφέροντος των πελατών για ένα προϊόν και ως εκ τούτου παραγωγή κερδοφόρων αποτελεσμάτων)
- Ανάλογα με την ανωτέρω ανάλυση, θα πρέπει να επιλεγθεί το σωστό μέσο διαφήμισης και προώθησης για συγκεκριμένο πελάτη εντός του συνόλου.



5. Διάθεση χώρου στα ράφια σύμφωνα με την ευκαιρία (προοπτική)

- Πόσα αποθέματα μπορούν να υπάρχουν στο ράφι, καθώς και στο κατάστημα σαν ασφάλεια
- Πόσο χώρος να διατεθεί σε κάθε προϊόν και μάρκα
- Πώς να γίνει μία αναδιάταξη (ανασύσταση) ραφιού

Για όλες αυτές τις πληροφορίες, υπάρχει μία ομάδα που αποτελείται από σχεδιαστές καταστημάτων, αναλυτές, κατανεμητές εμπορευμάτων και χώρου.

Πλεονεκτήματα

1. Προγραμματισμός & σχεδιασμός: Είναι μία χρήσιμη ανάλυση για τον σχεδιασμό του καταστήματος, το μάρκετινγκ, τις συνδυαστικές προωθήσεις και τις ενέργειες μέσα στο κατάστημα (merchandising)
2. Ποικιλία: Η απόφαση για την καλύτερη γκάμα λαμβάνεται σε επίπεδο κατηγορίας και υποκατηγορίας
3. Κατανομή χώρων: Βοηθά στην μακρο και μικρο κατανομή του χώρου για την κατηγορία / το προϊόν / τη μάρκα
4. Η διαχείριση των αποθεμάτων: Βελτιστοποίηση αποθεμάτων, βελτίωση της διαθεσιμότητας, σχεδιασμός αναπλήρωσης
5. Διαχείριση εσόδων:
  - Προσαρμοσμένες προωθητικές ενέργειες στις ειδικές ανάγκες κάθε ομάδας
  - Αύξηση των πωλήσεων με τον εντοπισμό ευκαιριών πωλήσεων
6. Τεράστια γκάμα προϊόντων:
  - Παρέχει στους πελάτες μεγάλη ποικιλία επιλογών
  - Παρέχει τα θεμέλια για να σχεδιαστεί ένα ποικιλόμορφο περιβάλλον πολυτοποθέτησης κατά τρόπο αποτελεσματικό και όχι χρονοβόρο
7. Εκχώρηση / ανακατανομή των καταστημάτων στις ομάδες: Νέα καταστήματα μπορούν εύκολα να ταξινομηθούν σε μια συστάδα ώστε να καθιερωθούν και να αναπτυχθούν γρήγορα. Παλιότερα καταστήματα μπορούν να εκχωρηθούν εκ νέου σε άλλες ομάδες αν αλλάξουν τα μοτίβα πωλήσεων ή αν γίνουν αλλαγές στην αγορά.

#### 4.5.1 Ομαδοποίηση με βάση τις αγοραστικές περιστάσεις

Μια εναλλακτική προσέγγιση είναι να ομαδοποιηθούν καταστήματα με βάση τη συμμετοχή της κάθε αγοραστικής περίπτωσης σε κάθε κατάσταση. Η κύρια διαφορά

είναι ότι στην περίπτωση της ομαδοποίησης καταστημάτων ενδέχεται να υπάρξουν πρόσθετες μετρήσεις με βάση τα κοινωνικο-δημογραφικά ή τοπικά χαρακτηριστικά που έχει κάθε κατάσταση, όπως το επίπεδο εισοδήματος των πελατών και της επιφάνειας του καταστήματος.

#### 4.6 Store Group Profilers

Η συγκεκριμένη λύση προσφέρεται για τους οργανισμούς που θέλουν να τρέξουν πιλοτικές δοκιμές σε μία ομάδα καταστημάτων, ώστε να δουν κατά πόσο αποδοτική είναι μία ενέργεια (όπως ένα καινούργιο προϊόν, μία καινούργια συσκευασία, μία καινούργια γεύση, μία αλλαγή τιμής, μία αλλαγή κωδικολογίου, μία αλλαγή στην επικοινωνία ή στις ενέργειες που γίνονται μέσα στα καταστήματα). Και αφού ποσοτικοποιηθούν οι συνέπειες αυτής της ενέργειας, στη συνέχεια να την εφαρμόσουν σε ολόκληρη την αγορά. Τα Store Group Profilers συμβάλλουν στο να εκμεταλλευτούν ευκαιρίες και να εφαρμοστούν οι καλύτερες πρακτικές ή να βελτιωθούν οι ήδη υπάρχουσες με σημαντικά χαμηλότερο κόστος και κίνδυνο.

Δημιουργούνται ομάδες καταστημάτων βάσει στατιστικής μεθοδολογίας στις οποίες θα εφαρμοστεί η ενέργεια και ομάδες καταστημάτων που θα χρησιμοποιηθούν για έλεγχο. Τα καταστήματα ελέγχου είναι παρόμοια με τα καταστήματα στα οποία θα γίνει η δοκιμή. Τα καταστήματα αντιστοιχίζονται με βάση διάφορα χαρακτηριστικά όπως τετραγωνικά μέτρα, πωλήσεις της μάρκας και της κατηγορίας, τάσεις πωλήσεων κατηγορίας και προϊόντος, κωδικολόγιο. Στη συνέχεια γίνεται στατιστική ανάλυση ώστε να μετρηθεί το καθαρό αποτέλεσμα της μεταβλητής που εξετάζεται. Συνήθως χρειάζονται στοιχεία για τις 52 εβδομάδες πριν τρέξει η ενέργεια, στοιχεία 4 εβδομάδων που τρέχει η ενέργεια και στοιχεία 4 εβδομάδων μετά το πέρας της ενέργειας. Έτσι, μετράται το αποτέλεσμα μίας νέας ενέργειας σε όρους επιπλέον όγκου, επιπλέον αξίας και αντίστοιχα των μεριδίων αγοράς.

Η ανάλυση Store Group Profilers βοηθάει στην δημιουργία ενός πειστικού σεναρίου για τους βιομήχανους στους λιανέμπορους, να τοποθετήσουν για παράδειγμα ένα νέο προϊόν στα καταστήματά τους δείχνοντάς τους πόσο παραπάνω θα ανεβούν οι πωλήσεις της κατηγορίας από το συγκεκριμένο προϊόν. Συμβάλλει στην αποφυγή κοστοβόρων λαθών από λανθασμένες αλλαγές στην τιμολογιακή πολιτική. Επιβεβαιώνει το βέλτιστο κωδικολόγιο για την καλύτερη απόδοση μίας επένδυσης

(ROI). Βοηθάει στην απόκτηση γρήγορης και ακριβούς πληροφόρησης που χρειάζεται για να υποστηριχθούν σημαντικές επιχειρηματικές αποφάσεις. Διασφαλίζει τις καλύτερες συνθήκες αποδοτικότητας σε πωλήσεις ενός προϊόντος.

Τα Store Group Profilers, είναι μία αρκετά ευέλικτη λύση που μπορεί να συμπεριλάβει οποιαδήποτε εξεζητημένη περίπτωση της αγοράς που πιθανό να μην μπορεί να μετρηθεί και να δώσει το καθαρό αποτέλεσμα μίας συγκεκριμένης μεταβλητής μάρκετινγκ. Μπορούν να συνδυαστούν με έρευνες αγοραστικής συμπεριφοράς, με τη χρήση δομημένου ερωτηματολογίου μέσα στα καταστήματα και πέρα από τα αποτελέσματα των δεδομένων, να υπάρχει και η άποψη του αγοραστή για ένα νέο προϊόν ή μία καινούργια γεύση ή για μία νέα επικοινωνία που αφορά μία συγκεκριμένη προσφορά.

#### 4.7 Μοντέλα προβλέψεων (Forecasting)

Η αβεβαιότητα για το μέλλον και οι αλλαγές στις συνθήκες του περιβάλλοντος που συμβαίνουν όλο και πιο έντονα τον τελευταίο καιρό, η καταναλωτική εμπιστοσύνη καθώς και η πιστότητα των αγοραστών που μειώνονται χρόνο με τον χρόνο λόγω του οικονομικού περιβάλλοντος, δημιουργούν όλο και πιο έντονα την ανάγκη για τις επιχειρήσεις να δουν την προοπτική τους στο μέλλον. Θέλουν να δουν την επίδραση που θα έχουν οι πωλήσεις τους από την κρίση, πως θα κάνουν πλάνα, ποιους στόχους θα βάλουν βάσει κάποιων δεδομένων και πως θα δημιουργήσουν μία στρατηγική ανάπτυξης μέσα σε ένα περιβάλλον γεμάτο προκλήσεις. Οι σωστές προβλέψεις είναι σημαντικές για την επίτευξη τόσο των στρατηγικών όσο και των λειτουργικών στόχων.

Η συγκεκριμένη λύση εντοπίζει εκείνα τα σημεία που συμβάλλουν στην ανάπτυξη της μάρκας και της κατηγορίας σε χρονικό ορίζοντα πέντε ετών. Το αποτέλεσμα του αναλυτικού μοντέλου ποσοτικοποιεί τις συνέπειες των σημείων αυτών και παρέχει μία πρόβλεψη πωλήσεων τόσο σε όγκο όσο και σε αξία. Παράλληλα, λαμβάνει υπόψη της ένα μεγάλο αριθμό παραγόντων όπως οικονομικούς δείκτες, δημογραφικές αλλαγές, κοινωνικοοικονομικές και καταναλωτικές τάσεις.

Τα οφέλη για τις βιομηχανίες ή τους λιανέμπορους είναι η καλύτερη κατανομή των πόρων ανάλογα με τους παράγοντες που επηρεάζουν την αύξηση των πωλήσεων μίας μάρκας ή μίας κατηγορίας. Κατά συνέπεια, η αύξηση της παραγωγικής και λειτουργικής

αποτελεσματικότητας, η μεγαλύτερη σταθερότητα στη δημιουργία και εκτέλεση των πλάνων. Οι βιομήχανοι και οι λιανέμποροι μπορούν να κρατούν τον ιδανικό αριθμό αποθήκευσης εμπορευμάτων ανάλογα με τις πωλήσεις που προβλέπονται για την κάθε χρονιά. Έχουν μεγαλύτερη ευελιξία για την ανταπόκριση στις προτιμήσεις των καταναλωτών, την αύξηση της κερδοφορίας και της απόδοσης μίας επένδυσης (ROI) μέσω των πλάνων που στηρίζονται σε δεδομένα και παράλληλα έχουν μικρότερο κίνδυνο για μείωση των πωλήσεων μέσω έγκαιρων και πληροφορημένων προσαρμογών των προωθητικών ενεργειών. Η συγκεκριμένη λύση δίνει τη δυνατότητα για δημιουργία σεναρίων μέσω ενός εργαλείου όπου μπορούν να γίνουν αλλαγές στις μεταβλητές, για παράδειγμα αύξηση κατά 10% της προωθητικής δραστηριότητας τον επόμενο χρόνο και να παίρνουν το αποτέλεσμα σε όγκο και αξία.

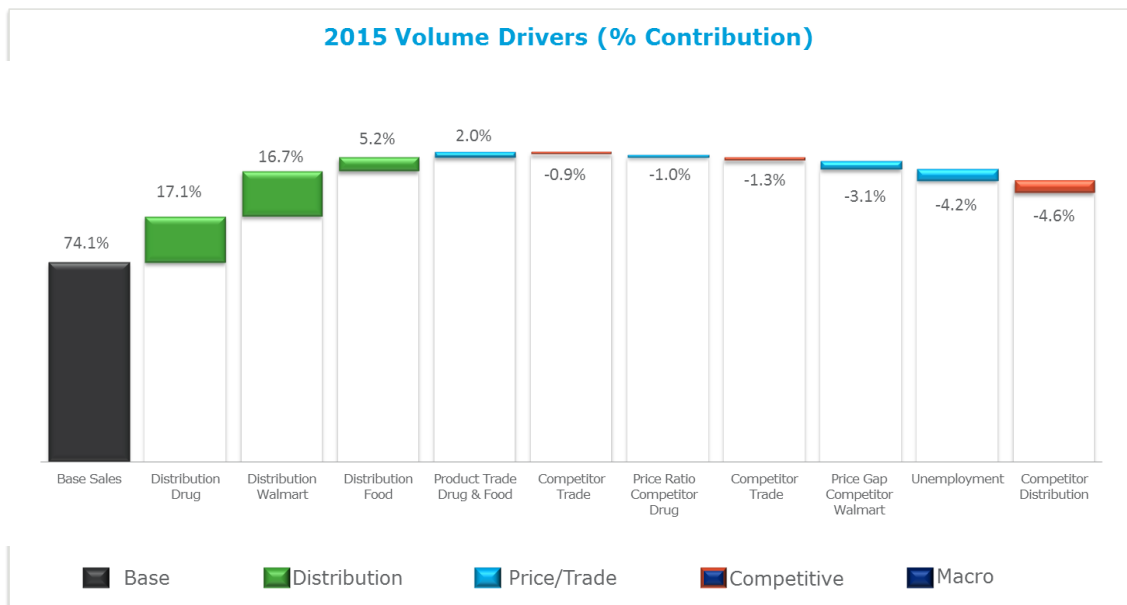
Παράλληλα, η συγκεκριμένη λύση επιτρέπει καλύτερη στρατηγική σε συνδυασμό με τους λιανέμπορους, για παράδειγμα σε ποιες λιανεμπορικές αλυσίδες θα πρέπει να στοχεύσει ο βιομήχανος τον επόμενο χρόνο ώστε να πετύχει τους στόχους του σε πωλήσεις, σε αξία, σε μερίδιο αγοράς, σε κερδοφορία. Επιπλέον, επιτρέπει την πιο αποτελεσματική τοποθέτηση και επικοινωνία της μάρκας μέσα από την ολοκληρωμένη κατανόηση των παραγόντων που επηρεάζουν την αγοραστική συμπεριφορά. Οι μαρκετίερς πλέον ξέρουν σε ποια στοιχεία του μίγματος μάρκετινγκ θα πρέπει να επενδύσουν ώστε να πετύχουν τους στόχους τους. Ακόμη, συμβάλλει στην μακροπρόθεσμη αύξηση της κερδοφορίας μέσω της κατανόησης των εξωτερικών παραγόντων που επηρεάζουν τον οργανισμό του, όπως το οικονομικό περιβάλλον, οι δημογραφικές αλλαγές. Για παράδειγμα, αν ένας βιομήχανος που παρασκευάζει γιαούρτι, είναι σε μία χώρα που ο αριθμός των ηλικιωμένων αυξάνεται με πολύ μεγάλο ρυθμό, πιθανό να προκύψει ανάγκη να δημιουργήσει γιαούρτι εμπλουτισμένο σε ασβέστιο.

Τα αναλυτικά μοντέλα προβλέψεων παράγουν πρόβλεψη ανά εβδομάδα, ανά τετραβδόμαδο, ανά τρίμηνο και δείχνουν ποιοι είναι οι παράγοντες που επηρεάζουν κατά κύριο λόγο τις πωλήσεις μίας μάρκας ή μίας κατηγορίας και ποσό. Παρακάτω ακολουθούν δύο παραδείγματα του παραδοτέου υλικού:

Πίνακας 2: Παράδειγμα παραδοτέου υλικού Μοντέλου Προβλέψεων (Forecasting)

Πηγή: IRI Analytics

Driver Type	Driver	% Contribution to Category	
Base	Base	68.7%	68.7%
Items per Store	Avg Items per Store	13.1%	13.1%
Price /Trade	Wtd Avg % PrcRed, Feature Only	4.8%	9.1%
	% ACV, Feature Only	2.9%	
	Wtd Avg % PrcRed, Display Only	1.5%	
Macroeconomic	Personal Income	2.5%	4.9%
	Consumer confidence Index	2.3%	
Seasonality	Seasonality	1.9%	2.0%
	Christmas	0.1%	
Distribution	# Stores Selling_Total US	1.3%	1.3%
Media	Total TV GRP	0.8%	0.9%
	Hispanic TV	0.1%	



Διάγραμμα 1: Παράδειγμα παραδοτέου υλικού Μοντέλου Προβλέψεων (Forecasting)

Πηγή: IRI Analytics

Τα στοιχεία που μπαίνουν στο μοντέλο είναι στοιχεία EPOS τουλάχιστον τριών ετών και οποιαδήποτε σχετική πληροφορία όπως θερμοκρασίες, χτυπήματα στο google, μακροοικονομικοί δείκτες (όπως ΑΕΠ, ανεργία, πληθωρισμός, δείκτες καταναλωτικής εμπιστοσύνης, τιμές καυσίμων), media όπως τηλεόραση, τύπος, GRPs και αντίστοιχα κόστη. Ακόμη μπορούν να μπουν αποτελέσματα από καταναλωτικές έρευνες, τάσεις καινοτομίας στην κατηγορία, εποχικότητα. Στη συνέχεια μοντέλα παλινδρόμησης εκτιμούν τους βασικούς παράγοντες που επηρεάζουν τις διακυμάνσεις των πωλήσεων καθώς και υποκείμενες τάσεις.

## 4.8 Αγοραστικές περιστάσεις (Shopper Occasions)

Ένας καταναλωτής επισκέπτεται ένα κατάστημα για περισσότερες από μία περιστάσεις. Για παράδειγμα, μπορεί να πάει τη μία φορά και να αγοράσει ό, τι ένα νοικοκυριό χρειάζεται, να κάνει δηλαδή τα ψώνια ρουτίνας και μια άλλη φορά θα μπορούσε να αγοράσει φρέσκο γάλα, που χρειάζεται κάθε 2 ή 3 ημέρες. Κάθε διαφορετική περίσταση είναι μία αγοραστική περίσταση και αντιπροσωπεύει έναν διαφορετικό λόγο για τον καταναλωτή που πηγαίνει να αγοράσει τα προϊόντα. Αγοραστικές περιστάσεις - χωρίς τη χρήση προγράμματος επιβράβευσης – είναι ο εντοπισμός των προτύπων συμπεριφοράς σε διαφορετικά σύνολα καλαθιών και η κατανόηση των διαφόρων παραγόντων που οδηγούν στην κάθε συμπεριφορά. Σε γενικές γραμμές μία αγοραστική περίσταση είναι απλά ένα σύνολο από καλάθια που έχουν περισσότερες ομοιότητες παρά διαφορές. Υπάρχουν διάφοροι τρόποι για να καθοριστούν ποια καλάθια ανήκουν σε ποιες αγοραστικές περιστάσεις. Μέχρι στιγμής, υπάρχουν τρεις διαφορετικές δυνατότητες:

1. Χρησιμοποιώντας επιχειρηματική λογική/επιχειρηματικούς κανόνες
2. Με τη χρήση της μεθόδου ανάλυσης συστάδων/με βάση τα πρότυπα κατανάλωσης
3. Με την ταξινόμηση νέων καλαθιών σε υφιστάμενες περιστάσεις

Για να αποφευχθεί η σύγχυση, χρησιμοποιούνται δύο διαφορετικοί όροι για να περιγραφεί η ίδια έννοια σε διαφορετικά επίπεδα, το αγοραστικό ταξίδι και η αγοραστική περίσταση. Αν δύο αγοραστικές περιστάσεις και το αγοραστικό ταξίδι παραπέμπουν κατ 'ουσίαν σε μια σειρά από καλάθια, το αγοραστικό ταξίδι θα υπερισχύσει των αγοραστικών περιστάσεων, καθώς ένα αγοραστικό ταξίδι μπορεί να περιλαμβάνει πολλές αγοραστικές περιστάσεις.

### 4.8.1 Ομαδοποίηση με τη χρήση επιχειρηματικών κανόνων

Η κάθε περίπτωση του πελάτη μπορεί να οδηγήσει σε διαφορετικές κατηγορίες καλαθιών που μπορεί να παρουσιάζουν ιδιαίτερο ενδιαφέρον. Για παράδειγμα, ένας λιανέμπορος θα μπορούσε να έχει συμφέρον για τον εντοπισμό των διαφορών μεταξύ των καλαθιών ανά ημέρα της εβδομάδας ή ακόμα και σε ώρες της ημέρας. Επίσης, μπορεί να είναι ενδιαφέρον να εξεταστούν καλάθια που έχουν ορισμένους τύπους

προϊόντων, δηλαδή φαγητό διαίτης. Ή μπορεί κάποιος να θέλει να δει τις διαφορές ανάμεσα στα μεγάλα και στα μικρά καλάθια ή ανάμεσα στα καλάθια με και χωρίς προσφορές. Κατά παρόμοιο τρόπο μία βιομηχανία βρεφικών προϊόντων θα ενδιαφερόταν για τα καλάθια που περιλαμβάνουν βρεφικά προϊόντα και περισσότερο για τα καλάθια με τα δικά της προϊόντα. Αυτές οι κατηγορίες προέρχονται συνήθως από συγκεκριμένες επιχειρηματικές ερωτήσεις ή από περιπτώσεις που θα μπορούσαν να περιλαμβάνουν ενδιαφέρουσες ιδέες.

#### 4.8.2 Ομαδοποίηση με βάση τα πρότυπα κατανάλωσης

Σε αντίθεση με την περίπτωση που περιγράφεται παραπάνω, συνήθως ο πελάτης έχει μια σειρά από βασικούς μοχλούς που είναι σημαντικοί επιχειρηματικά αλλά ακόμη δεν υπάρχει καμία γνώση σχετικά με την αριθμητική σημασία αυτών των μεταβλητών, καθώς και των διαφορετικών συνδυασμών τους ή του μεγέθους των συνδυασμών τους. Αυτό σημαίνει ότι μία καταμέτρηση των κριτηρίων που προτείνονται από τον πελάτη μπορεί να οδηγήσει σε μια πολύ μικρή ομάδα καλάθιων που στο τέλος έχουν ελάχιστη ή και καθόλου σημασία σε ολόκληρη την επιχείρηση.

Για να αντιμετωπιστεί αυτό, γίνεται ομαδοποίηση των καλάθιων, χρησιμοποιώντας τους βασικούς μοχλούς που καθοδηγούν τον πελάτη. Με αυτόν τον τρόπο, δημιουργούνται διαφορετικές ομάδες καλάθιων με βάση τις βασικές κινητήριες δυνάμεις, όμως τα κριτήρια αυτά βασίζονται τόσο στις ομοιότητες των καλάθιων που ανήκουν στην ίδια ομάδα όσο και στις διαφορές μεταξύ καλάθιων που ανήκουν σε διαφορετικές ομάδες. Αυτό δεν σημαίνει ότι θα υπάρχει ένας εξαντλητικός κατάλογος των συνδυασμών των χαρακτηριστικών, αλλά μάλλον ένα δυναμικό σύνολο συνδυασμού ιδιοτήτων με βάση τις αριθμητικές πληροφορίες στο σύνολο δεδομένων.

#### 4.8.3 Ταξινόμηση νέων καλάθιων σε υπάρχουσες περιπτώσεις

Η περίπτωση αυτή ισχύει όταν ένα σύνολο αγοραστικών περιστάσεων έχει ήδη δημιουργηθεί και νέα καλάθια κατανέμονται σε κάθε μία από αυτές τις περιστάσεις. Συνήθως, η αγοραστική περίσταση έχει συμπληρωθεί με βάση τα δεδομένα του παρελθόντος και νέα δεδομένα κατανέμονται σε αυτά για τον σκοπό της "συνέχειας". Δεδομένου ότι η ομαδοποίηση δεν είναι μια ακριβής επιστήμη, δεν υπάρχει γνωστός

τρόπος για να αναπαραχθεί το περιεχόμενο των συνεργατικών σχηματισμών. Στην πράξη, ακόμη και μια μικρή αλλαγή στο περιεχόμενο του καλαθιού μπορεί να οδηγήσει σε εντελώς διαφορετικές ομάδες.

## 4.9 Δίκτυα προϊόντων (Product Networks)

### 4.9.1 Τι είναι ένα δίκτυο προϊόντων

Δίκτυο προϊόντων είναι ένα γράφημα. Κάθε κόμβος (κορυφή) στο δίκτυο αντιπροσωπεύει ένα προϊόν / κατηγορία προϊόντος, όπως μία κύρια ομάδα, κατηγορία κλπ, και μία ακμή (γραμμή) εμφανίζεται μεταξύ δύο οποιωνδήποτε κόμβων που έχουν αγοραστεί μαζί σε μια συναλλαγή. Γενικά, το μέσο προϊόν δεν αγοράζεται συχνά με την πλειοψηφία των προϊόντων που γειτνιάζει. Αγοράζεται συχνά μόνο με μερικά από αυτά.

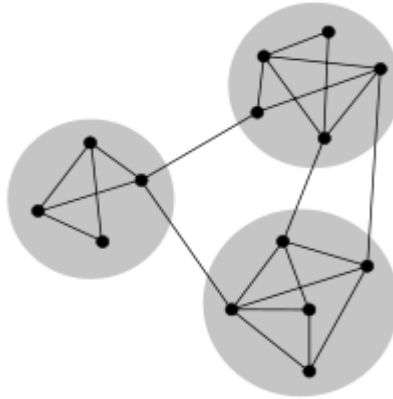
Διαφέρουν από τους άλλους τύπους δικτύων για έναν απλό λόγο: η παρουσία μιας ακμής δεν συνεπάγεται κατ'ανάγκη την επιβεβαίωση της σχέσης μεταξύ των προϊόντων. Για παράδειγμα, αν ένα άτομο αγοράζει χαρτί κουζίνας και σάλτσα για μακαρόνια στην ίδια συναλλαγή, αυτό δεν συνεπάγεται ένα κοινό κίνητρο για τις δύο αγορές. Από την άλλη πλευρά, ένα πρόσωπο που αγοράζει πολλά ανεξάρτητα προϊόντα σε μια ενιαία συναλλαγή θα μπορούσαν να αποτελούν μια κλίκα μεταξύ τους, παρά την απουσία πραγματικής σχέσης. Ως αποτέλεσμα, τα δίκτυα προϊόντων είναι πολύ πυκνά, με ένα μεγάλο αριθμό συνδέσεων ανά κόμβο. Ωστόσο, πολλές από αυτές τις άκρες είναι χωρίς νόημα: αντιπροσωπεύουν πλαστές ενώσεις που παράγονται τυχαία. Από τη άλλη πλευρά, θα υπάρξουν εξαιρετικά άκρα χαμηλής σημαντικότητας που πιθανό να αντιπροσωπεύουν ισχυρές σχέσεις.

### 4.9.2 Δομή σε ένα δίκτυο προϊόντων

Στη μελέτη των δικτύων, έχει βρεθεί ένας αριθμός διαφορετικών χαρακτηριστικών που υπάρχουν συχνά. Ένας από αυτούς είναι η δομή της κοινότητας. Η δομή της κοινότητας αναφέρεται στην εμφάνιση ομάδων κόμβων (προϊόντα / ομάδες προϊόντων)



σε ένα δίκτυο που είναι περισσότερο πυκνά συνδεδεμένο με το υπόλοιπο του δικτύου (σχήμα 2). Αυτή η ομοιογένεια των συνδέσεων δείχνει ότι το δίκτυο έχει ορισμένες φυσικές διαιρέσεις στο εσωτερικό του. Οι κοινότητες συχνά ορίζονται με βάση το ότι κάθε κόμβος έχει τεθεί σε μία και μόνο μία κοινότητα, όπως και στο Σχήμα 2.



Σχήμα 6: Ένα σχέδιο μικρού δικτύου που δείχνει τη δομή μίας κοινότητας με τρεις ομάδες.  
Πηγή: IRI Analytics

Πως όμως ανιχνεύεται μία κοινότητα και πως οι κοινότητες των προϊόντων μπορούν να χρησιμοποιηθούν για να αποκτηθούν οι απαραίτητες γνώσεις για τη συμπεριφορά των πελατών σε ένα κατάστημα. Η ανακάλυψη των κοινοτήτων (clusters) των προϊόντων στα δεδομένα μπορεί να χρειάζεται περαιτέρω ανάλυση. Η προσέγγιση με βάση το δίκτυο μπορεί να απομονώσει την επιρροή μεταξύ των προϊόντων, μετριάζοντας την ανάγκη για αναζήτηση μέσα στις μαζικές λίστες των κανόνων συσχέτισης. Τέλος, είναι μια ολοκληρωμένη και καινοτόμα στρατηγική ανάλυση συνδυάζοντας τόσο παραδοσιακές τεχνικές όσο και τεχνικές που βασίζονται στα δίκτυα. Αυτό το πλαίσιο είναι σε θέση να παράγει ιδέες που είναι δύσκολο να βρεθούν με τις παραδοσιακές μεθόδους ανάλυσης. Τέλος, προτείνεται η ανίχνευση κοινότητας ως ένα πρώτο διερευνητικό βήμα στη διαδικασία ανάλυσης, όπου θα φωτίσει τις σχέσεις μεταξύ των σημαντικών, αλλά πιο απομακρυσμένων προϊόντων. Στη συνέχεια, η επακόλουθη ανάλυση των κανόνων συσχέτισης μπορεί να επικεντρωθεί περισσότερο εντατικά στα προϊόντα των οποίων ο ρόλος δεν είναι σαφής μέσα από την αποσύνθεση της κοινότητας, ίσως με τεχνικές όπως είναι οι κανόνες σύνδεσης δικτύων.

Για να αφαιρεθούν κάποιες από τις θορυβώδεις άκρες που δημιουργούνται από τυχαίες αγορές, όπως έχει ήδη αναφερθεί, και να υπάρξει βελτίωση της ποιότητας της παραπάνω ανάλυσης, πρέπει να δημιουργηθεί ένα ελάχιστο όριο μίας παραμέτρου.

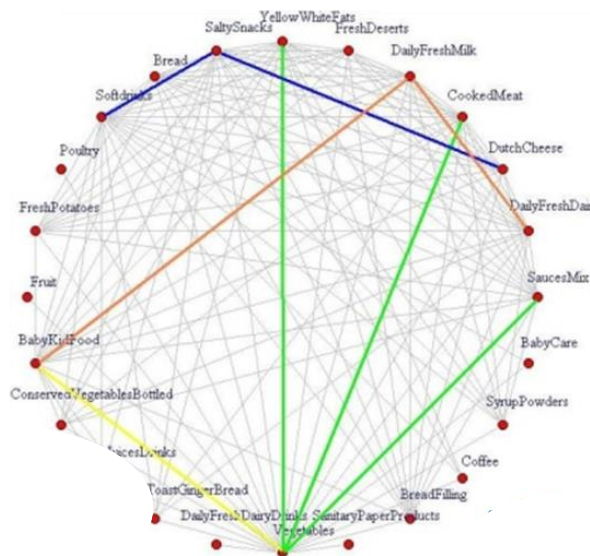
Αυτό σημαίνει ότι υπάρχει μια ακμή ανάμεσα σε δύο προϊόντα, μόνον εάν έχουν αγοραστεί από κοινού τουλάχιστον  $\sigma$  φορές. Αυτό είναι ανάλογο με την επιλογή ενός κατώτατου ορίου στήριξης για κανόνες συσχέτισης. Για κάθε σχέδιο, η παράμετρος  $\sigma$  πρέπει να καθορίζεται με βάση την προηγούμενη γνώση των πληροφοριών από τις συναλλαγές κάθε λιανέμπορου. Έτσι, συνδυασμοί με μικρότερο αριθμό από ό, τι  $\sigma$  θα πρέπει να αφαιρεθούν από την ανάλυση ανίχνευσης της κοινότητας.

Ο σκοπός για τη δημιουργία των παραπάνω αρχείων είναι να δημιουργηθεί το αρχείο εισόδου για τη λειτουργία του αλγορίθμου ανίχνευσης της κοινότητας του δικτύου προϊόντων. Το αρχείο εισόδου αποτελείται από υπόλοιπα ζεύγη κύριων ομάδων που έχουν ένα δείκτη βαρύτητας. Από τα δύο προηγούμενα αρχεία, υπολογίζουμε τον Δείκτη Jaccard (ένα δείκτη βαρύτητας, οποιοσδήποτε άλλος μπορεί να χρησιμοποιηθεί) για κάθε ζεύγος των κύριων ομάδων. Ο δείκτης υπολογίζεται για κάθε ζεύγος και το αρχείο εισόδου για τον αλγόριθμο ανίχνευσης της κοινότητας είναι έτοιμο.

#### 4.10 Κανόνες συσχέτισης (Association Rules)

##### Σκοπός και Στόχος

Οι κανόνες συσχέτισης (AR) εμπίπτουν στην Άντληση Δεδομένων (Data Mining) για την ανακάλυψη των σχέσεων σε μεγάλα σύνολα δεδομένων, όπως "Οι πελάτες που παραγγέλνουν το προϊόν Α συχνά επίσης μπορεί να παραγγείλουν το προϊόν Β ή Γ". Κατά τους εν λόγω κανόνες συσχέτισης, μπορούν να ληφθούν αποφάσεις για αποτελεσματικό μάρκετινγκ που αφορούν τη διαφήμιση, την τιμολόγηση, την τοποθέτηση προϊόντων, την εισαγωγή νέας κατηγορίας, την στοχευμένη προβολή, τον σχεδιασμό καταλόγου, τη βέλτιστη συλλογή, τη διαρρύθμιση καταστήματος, ειδικές προσφορές με τους κατασκευαστές, κ.λπ. Σε γενικές γραμμές, ο κύριος στόχος των κανόνων συσχέτισης είναι να αναλύσουν και να περιγράψουν τα πρότυπα συμπεριφοράς των πελατών, όπως φαίνεται στο παρακάτω σχήμα 3.



Σχήμα 7: Διαφορετικές ομάδες κανόνων συσχέτισης μπορεί να υποκρύπτουν την ύπαρξη ομάδων καταναλωτών που έχουν διαφορετικές συμπεριφορές.

Πηγή: IRI Analytics

#### Επίσημος ορισμός και χαρακτηριστικά

Προκειμένου να μετρηθεί η σημαντικότητα ενός κανόνα συσχέτισης χρησιμοποιούνται 3 βασικές μετρήσεις, η υποστήριξη, η εμπιστοσύνη και η ανύψωση.

- Η υποστήριξη δείχνει πόσο συχνά δύο (ή περισσότερα) προϊόντα εμφανίζονται στην ίδια απόδειξη  $P(A \cap B)$

- Η εμπιστοσύνη μετρά την πιθανότητα να αγοραστεί το προϊόν A, με δεδομένο το γεγονός ότι το προϊόν B έχει ήδη αγοραστεί και το αντίστροφο.  $P(A | B) = \frac{P(A \cap B)}{P(B)}$ ,

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

- Η ανύψωση (Lift) μετρά την απόδοση ενός μοντέλου στην πρόβλεψη ή την ταξινόμηση των υποθέσεων. Είναι ο λόγος της παρατηρούμενης στήριξης προς την αναμενόμενη αν τα X και Y ήταν ανεξάρτητα. Αυτό το μέτρο είναι ένας λόγος, δεν είναι μια πιθανότητα:

$$Lift = \frac{P(A | B)}{P(A)} = \frac{\frac{P(A \cap B)}{P(B)}}{P(A)} = \frac{P(A \cap B)}{P(A) * P(B)}$$

Η πρώτη μέτρηση που πρέπει να γίνει είναι αυτή της παραμέτρου της ανύψωσης (Lift). Όταν η ανύψωση ενός κανόνα είναι (περισσότερο ή λιγότερο) ίση με 1, αυτό σημαίνει ότι η συχνότητα με την οποία ένας δεδομένος συνδυασμός εμφανίζεται είναι παρόμοια

με την αναμενόμενη συχνότητα της παραδοχής ότι οι πωλήσεις των προϊόντων στον συνδυασμό που μελετήθηκε είναι ανεξάρτητες μεταξύ τους. Αυτό σημαίνει ότι αυτός ο κανόνας δεν είναι στατιστικά σημαντικός. Οι τιμές, που είναι υψηλότερες από 1, σημαίνει ότι ο συνδυασμός παρατηρείται συχνότερα από το αναμενόμενο και επομένως ότι αυτά τα προϊόντα πράγματι συχνά αγοράζονται μαζί. Τιμές μικρότερες από 1 από την άλλη πλευρά, συνεπάγονται ότι ο συνδυασμός παρατηρείται λιγότερο συχνά από το αναμενόμενο, και ως εκ τούτου αυτά τα προϊόντα δεν αγοράζονται συχνά μαζί, παρόλο που δεν είναι δόκιμος ο όρος, αυτός ο κανόνας ονομάζεται και «ποτέ δεν αγοράζονται μαζί». Από την άποψη αυτή, η ανύψωση χρησιμοποιείται για να βρεθούν οι κανόνες που έχουν ενδιαφέρον και να γίνει διάκριση μεταξύ των κανόνων «συχνά αγοράζονται μαζί» και «ποτέ δεν αγοράζονται μαζί».

Μετά από αυτό, λαμβάνει χώρα η μέτρηση της υποστήριξης, ανάλογα με το μέγεθος του συνόλου δεδομένων και το πεδίο εφαρμογής. Οι κανόνες πρέπει να έχουν τόσο νόημα, όσο και η σημασία τους. Αυτό σημαίνει ότι πρέπει να ισχύουν για ένα σημαντικό αριθμό περιπτώσεων, ώστε να έχουν πραγματικό αντίκτυπο ή με άλλα λόγια, να έχουν επαρκή υποστήριξη. Ας υπενθυμίσουμε ότι η υποστήριξη είναι στην ουσία το ποσοστό επί τοις εκατό των καλαθιών που έχουν ένα συνδυασμό προϊόντων σε ολόκληρο το σύνολο δεδομένων από καλάθια.

Η εμπιστοσύνη είναι η ουσία των κανόνων συσχέτισης. Μεγάλη πιθανότητα για την εμπιστοσύνη δείχνει ότι όταν το πρώτο μέρος του κανόνα συμβαίνει, τότε είναι πολύ πιθανό, να συμβεί και το δεύτερο μέρος του κανόνα. Η εμπιστοσύνη, σε αντίθεση με την υποστήριξη και την ανύψωση, είναι διαφορετική όταν υπάρχουν αλλαγές κατάστασης. Για το λόγο αυτό πρέπει να δοθεί ιδιαίτερη προσοχή στο πρώτο μέρος και το δεύτερο μέρος του κάθε κανόνα. Έτσι, η υποστήριξη πρέπει να είναι μεγάλη, αλλά μπορούν να γίνουν και συμβιβασμοί. Εάν το σύνολο δεδομένων είναι μεγάλο, τότε θα μπορούσε να μειωθεί το κατώτατο όριο για την υποστήριξη. Τέλος, οι κανόνες θεωρούνται επίσης σύμφωνα με την πιθανή χρήση τους. Για παράδειγμα, αν ένας λιανέμπορος επικεντρώνεται στους συσχετισμούς μεταξύ των βρεφικών προϊόντων, τότε η εμπιστοσύνη και η υποστήριξη μπορούν να είναι μικρότερες, προκειμένου οι απαιτούμενες διατάξεις να αναδυθούν. Είναι μια κοινή πρακτική σε τέτοιες περιπτώσεις, να περιορίζεται το σύνολο δεδομένων (π.χ. αποδείξεις μόνο με βρεφικά προϊόντα) σε ένα υποσύνολο το οποίο είναι σημαντικό για την συγκεκριμένη κατηγορία. Με αυτόν τον τρόπο οι κανόνες για την κατηγορία που διερευνάται (δηλαδή βρεφικά προϊόντα) θα γίνονται ισχυρότεροι και αισθητοί.

Σε γενικές γραμμές ένας κανόνας συσχέτισης μπορεί να συνδέσει έναν αριθμό προϊόντων στη δεξιά πλευρά με μια σειρά προϊόντων στην αριστερή πλευρά, δηλαδή οι άνθρωποι που αγοράζουν τα προϊόντα Α, Β και Γ, συνήθως αγοράζουν από κοινού και προϊόντα D και E. Στην πράξη, αν ένας κανόνας σύνδεσης είτε συνδέει δύο προϊόντα - «οι άνθρωποι που αγοράζουν το προϊόν Α αγοράζουν συνήθως μαζί και προϊόντα D» - είτε συνδέει μια σειρά προϊόντων σε ένα προϊόν – «οι άνθρωποι που αγοράζουν τα προϊόντα Α, Β και Γ, συνήθως αγοράζουν από κοινού επίσης προϊόν Ε». Στην πρώτη περίπτωση χρησιμοποιούνται ζεύγη κανόνα και στην δεύτερη περίπτωση κανόνες πλειάδας. Ένας κανόνας πλειάδα περιλαμβάνει περισσότερες πληροφορίες αλλά είναι πιο δύσκολο να παραχθεί και να διαχειριστεί, ακόμη και να απεικονιστεί. Κατά συνέπεια, συχνά προτιμούνται οι κανόνες για ζεύγη προϊόντων. Μια πλειάδα μπορεί να σπάσει σε μια σειρά κανόνων ζευγών. Για τους κανόνες συσχέτισης χρησιμοποιούνται δύο διαφορετικές προσεγγίσεις, μία για κάθε περίπτωση. Ειδικότερα για τους κανόνες πλειάδας, χρησιμοποιείται WEKA ενώ για τους κανόνες κατά ζεύγη σενάρια C ++ ή SAS ερωτήματα.

## 5. IRI - Analytics Center of Excellence (ACE) [19]

Σε αυτό το κεφάλαιο αναλύεται μία μελέτη περίπτωσης. Η μελέτη περίπτωσης χρησιμοποιείται γενικά για να εξεταστεί ένα θέμα σε βάθος και λεπτομέρεια. Σε μία μελέτη περίπτωσης, η υπόθεση που μελετάται μπορεί να είναι ένα άτομο, μία οργάνωση, μία εκδήλωση ή μία ενέργεια που υπάρχει σε ένα συγκεκριμένο χώρο και χρόνο. Σύμφωνα με μια προσέγγιση, η μελέτη περίπτωσης αποτελεί ερευνητική στρατηγική, μια εμπειρική έρευνα που ερευνά ένα φαινόμενο στο πλαίσιο της πραγματικής ζωής του. Στη συγκεκριμένη περίπτωση, μελετάται η εξαγωγή υπηρεσιών υψηλού επιστημονικού επιπέδου, από την Ελλάδα στις υπόλοιπες χώρες του κόσμου, σε περίοδο οικονομικής κρίσης όπου μία πολυεθνική εταιρεία, η IRI, μεταφέρει στην Ελλάδα το παγκόσμιο αναλυτικό της κέντρο, το ACE (Analytics Center of Excellence).

Το ACE αποτελεί το παγκόσμιο αναλυτικό κέντρο της IRI και έχει έδρα στην Ελλάδα. Μέχρι σήμερα αριθμεί 150 εργαζομένους και αναπτύσσεται συνεχώς. Το αντικείμενό του είναι να δημιουργεί αναλυτικά μοντέλα για όλο τον κόσμο, κυρίως για εταιρείες που δραστηριοποιούνται στον τομέα των ταχυκίνητων καταναλωτικών προϊόντων. Το ACE στην Ελλάδα δημιουργήθηκε το 2009. Όταν όλοι αποεπένδυαν από μία αγορά και η οικονομία εισερχόταν στην περιδίνηση των μνημονίων και όλο και περισσότεροι νέοι Έλληνες έφευγαν στο εξωτερικό προς αναζήτηση εργασίας, η IRI και ο Hans Jessen, European Vice President, αποφάσισαν να μεταφέρουν το παγκόσμιο αναλυτικό κέντρο της εταιρείας από διάσπαρτα μικρότερα τμήματα στην Μπανγκαλόρ της Ινδίας, την Ευρώπη και τη Βόρεια Αμερική στην Αθήνα. Έτσι έγινε μία από τις σημαντικότερες άμεσες επενδύσεις (ίσως η πιο μεγάλη από πλευράς πρόσληψης επιστημονικά καταρτισμένου προσωπικού) – σε μία δύσκολη εποχή για την Ελλάδα.

Την στιγμή που μεγάλες πολυεθνικές επέλεξαν την Ινδία για λόγους τεράστιου ανθρώπινου κεφαλαίου και εξοικονόμησης πόρων και ντόπιοι επιχειρηματίες έπαιρναν τις εταιρείες τους και μετανάστευαν φορολογικά εκτός συνόρων, η IRI επενδύοντας στην έξοδο της Ελλάδας από την κρίση αποφάσισε -χωρίς να το πολυδιαφημίσει- να στεγάσει στην Αθήνα την πιο επίλεκτη ομάδα της εταιρείας στις αναλύσεις δεδομένων. Έτσι, το ACE (Analytics Center of Excellence), το οποία ξεκίνησε τα δύο πρώτα χρόνια με 40 προσλήψεις υψηλά καταρτισμένου επιστημονικού προσωπικού στη χώρα μας, σήμερα φτάνει να απασχολεί 150 υψηλά μορφωμένους Έλληνες επιστήμονες από τον τομέα των Μαθηματικών, του Μάρκετινγκ, της Ανάλυσης και της ευρύτερης αγοράς. Χρησιμοποιεί το καλύτερο δυνατό ανθρώπινο κεφάλαιο της χώρας και αξιοποιεί

επιστήμονες που υπό κανονικές συνθήκες θα είχαν ήδη μεταναστεύσει σε άλλες χώρες.

Ο βασικός λόγος για τον οποίο επιλέχθηκε η Ελλάδα για μια τόσο σημαντική επένδυση είναι το υψηλό διανοητικό κεφάλαιο των Ελλήνων σε συνδυασμό με το χαμηλό κόστος εργασίας. Όπως περιγράφει και ο Dr Jessen: «Ας είμαστε ξεκάθαροι. Οι Έλληνες έχετε ένα ισχυρό διανοητικό κεφάλαιο που σπάνια συναντιέται στον πλανήτη. Είστε αυτό που λέμε “independent thinkers”, ελεύθερα πνεύματα δηλαδή. Αυτό που ίσως σας λείπει είναι οι αυτοματισμοί, όμως σαφώς ο τρόπος σκέψης σας είναι σημαντικός και δεν μοιάζει με άλλων λαών. Αυτό λοιπόν που εκτιμήσαμε ως όμιλος -και επειδή πάντοτε το προσωπικό μας έχει υψηλά στάνταρτ- είναι το ότι οι Έλληνες ως εργαζόμενοι είναι άνθρωποι της πρωτοβουλίας. Είναι ευφάνταστοι και αποτελεσματικοί και σκέφτονται εκτός κανόνων, με την καλή έννοια. Η επένδυσή μας λοιπόν στη χώρα σας έχει να κάνει με δύο διαπιστώσεις: η Ελλάδα διαθέτει εξαιρετικά μαθηματικά μυαλά, πράγμα απαραίτητο για τις business μας, σε συνδυασμό με το δελεαστικό κόστος εργασίας, το οποίο δεν φτάνει φυσικά τόσο χαμηλά όσο στην Ινδία, όμως δεν αγγίζει τα όρια των ΗΠΑ ή της Βρετανίας. Αν μιλούσαμε, λοιπόν, για μια επένδυση που θα αναζητούσε χέρια εργασίας, αν αναφερόμασταν σε μια αμιγώς βιομηχανική επένδυση, θα μπορούσα να πω ότι η επιλογή της Ινδίας θα ήταν ενδεχομένως πιο συμφέρουσα. Επειδή όμως ο όμιλός μας βασίζεται στην ανάλυση, απαιτεί επιστημονική εργασία και χρειάζεται διανοητικά υψηλό κεφάλαιο, επέλεξε τους Έλληνες εργαζόμενους».

## 5.1 IRI & Analytics Center of Excellence (ACE)

Η IRI (Information Resources Inc.) είναι η 5η μεγαλύτερη εταιρεία ερευνών και μελετών αγοράς παγκοσμίως σύμφωνα με τα στοιχεία της American Marketing Association και απασχολεί 3.500 εργαζομένους σε όλο τον κόσμο. Ιδρύθηκε το 1979 στο Σικάγο των ΗΠΑ και σήμερα το πλειοψηφικό μερίδιο των μετοχών της ανήκει στη New Mountain Capital, ένα αμερικανικό fund που εστιάζει στην ανάπτυξη και όχι σε υψηλού κινδύνου ή μόχλευσης του χρέους επενδύσεις. Η IRI είναι ένας πολυεθνικός πάροχος υπηρεσιών και πληροφοριών επιχειρηματικής ευφυΐας και συμβουλευτικής, που υποστηρίζονται από πλατφόρμες τεχνολογίας, έμπειρες ομάδες ανθρώπινου δυναμικού και τη μοναδική της πρόσβαση σε πληροφορίες σχετικά με τις αγορές, τους καταναλωτές και αγοραστές με στόχο τη μετατροπή της «διορατικότητας» της IRI στις αγορές σε μετρήσιμα αποτελέσματα ανάπτυξης για τους πελάτες της. Είναι επίσης

ένας από τους κορυφαίους παγκόσμιους παρόχους «διανοητικού κεφαλαίου», καινοτόμων λύσεων και υπηρεσιών σχετικά με τις αγορές, τους καταναλωτές και τους αγοραστές. Δραστηριοποιείται σε 58 χώρες σε ολόκληρη την υφήλιο, από την Αυστραλία μέχρι τη Νότια Αμερική και από τη Ρωσία μέχρι τον Καναδά, μέσω των 11 γραφείων της στις ΗΠΑ και των 7 στην Ευρώπη, των θυγατρικών της, αλλά και ενός οικοσυστήματος συνεργασιών και εταιρικών συμμαχιών.

Στο πελατολόγιο της IRI περιλαμβάνεται το 95% των εταιρειών του δείκτη Fortune Global 500, δηλαδή του δείκτη των 500 μεγαλύτερων εταιρειών παγκοσμίως που δραστηριοποιούνται στον χώρο των ταχυκίνητων καταναλωτικών προϊόντων και του λιανεμπορίου. Συνολικά εξυπηρετεί περισσότερους από 1.000 μεγάλους πελάτες, όπως τους μεγαλύτερους παραγωγούς ταχυκίνητων καταναλωτικών προϊόντων και τις μεγαλύτερες αλυσίδες λιανεμπορίου. Χαρακτηριστικό στοιχείο της εταιρικής κουλτούρας της IRI είναι η επένδυση σε μια αμφίδρομη επωφελή σχέση με το πελατολόγιό της, κάτι που αποδεικνύεται από το γεγονός ότι ο μέσος όρος διάρκειας της συνεργασίας της με τους κυριότερους πελάτες της είναι περίπου 17 χρόνια.

Η IRI προσφέρει προϊόντα και υπηρεσίες σε τέσσερις τομείς: στις μετρήσεις αγοράς με παρακολούθηση δεδομένων πωλήσεων μέσω ενός εκτεταμένου δικτύου 95.000 σημείων άντλησης δεδομένων πωλήσεων σε Ευρώπη και Αμερική, στα Predictive Analytics, τις αναλυτικές τεχνικές πρόβλεψης των πωλήσεων, στα εργαλεία Λογισμικού Παρακολούθησης της Απόδοσης (Enterprise Performance Management Software) για την υποστήριξη ταχύτερης ανάλυσης του περιεχομένου της αγοράς και στις επαγγελματικές Υπηρεσίες / Συμβουλές με έμφαση σε τέσσερις βασικούς τομείς δραστηριότητας: βέλτιστες στρατηγικές τιμολόγησης, ανάπτυξη και καινοτομία, μάρκετινγκ καταναλωτικής συμπεριφοράς, καθώς και συμβουλευτική υποστήριξη στη διαχείριση των καναλιών διανομής. Από τα κυριότερα προϊόντα και υπηρεσίες της IRI είναι το InfoScan, όπου κάθε εβδομάδα αρκετές χιλιάδες σούπερ μάρκετ, πολυκαταστήματα, παντοπωλεία και φαρμακεία παραδίδουν τα στοιχεία των πωλήσεων των προϊόντων που συλλέγονται από τους σαρωτές. Επίσης υπάρχει το Consumer Network, που αποτελεί ένα παγκόσμιο πάνελ καταναλωτών και παρέχει στην IRI μια εσωτερική οπτική για τη συμπεριφορά των καταναλωτών από διάφορες πληθυσμιακές ομάδες νοικοκυριών.

Ακόμη υπάρχει η πλατφόρμα Liquid Data, που αναπτύχθηκε σε συνεργασία με την IBM και παρέχει άμεση πρόσβαση σε στρατηγικές πληροφορίες, προσφέρει δυνατότητες διαχείρισης και ενοποίησης δεδομένων από διαφορετικές πηγές, που επιτυγχάνεται



μέσω ενός συνδυασμού λογισμικών και κατοχυρωμένων με δίπλωμα ευρεσιτεχνίας αλγορίθμων της IBM και της IRI, και έχει τη δυνατότητα να αξιοποιεί την προγνωστική δυνατότητα των δεδομένων μέσα σε λίγα λεπτά. Τέλος, η εταιρεία διατηρεί και την υπηρεσία Thought Leadership: η IRI κυκλοφορεί τακτικά μια σειρά περιοδικών εκθέσεων που υπογραμμίζουν τις νέες εξελίξεις και τα κρίσιμα γεγονότα σε όλες τις σημαντικές προϊόντικές κατηγορίες.

## 5.2 ACE - Analytics Center of Excellence

Ο πραγματικός άσος στο μανίκι της IRI όμως είναι το ACE (Analytics Center of Excellence), το οποίο αποτελεί μια επιτελική επιχειρησιακή μονάδα της εταιρείας που ενσαρκώνει το όραμα του Δανού European Vice President, Dr Jessen να στεγάσει κάτω από μια κοινή επιχειρησιακή στέγη την επίλεκτη ομάδα της εταιρείας στα αναλυτικά μοντέλα. Η απόφαση του Dr Jessen να συγκεντρώσει την επίλεκτη ομάδα ανθρώπων στην Ελλάδα ήταν απολύτως συνειδητή, δεδομένης της ευχέρειας επιλογών να τοποθετήσει το ACE οπουδήποτε στον κόσμο έκρινε ο ίδιος.

Τον Ιούνιο του 2009 όμως εγκαθίσταται στην Ελλάδα και μέχρι το τέλος του 2010 αριθμεί κάτι λιγότερο από 40 εργαζομένους. Στη διάρκεια αυτού του πρώτου 1,5 χρόνου το ACE εξελίσσεται ραγδαία σε κρίσιμο κόμβο για τα Analytics της IRI στην Ευρώπη και συνεχίζει να αυξάνει τη διεξοδικότητά του στα έργα των ΗΠΑ. Μέχρι το τέλος του 2011 το σύνολο των αναλυτικών μοντέλων της Ευρώπης εκτελείται στο ACE, καθώς και το 50% των αντίστοιχων αμερικανικών μελετών της εταιρείας. Στη διάρκεια του 2011 ο αριθμός των εργαζομένων αυξάνεται κατά 50%, και στις αρχές του 2012 αγγίζει τους 60, στις αρχές του 2013 ο αριθμός αυτός ανέρχεται σε 74, ενώ τον Ιανουάριο του 2014 στους 88. Το 2014 είναι κομβικό έτος για το ACE, καθώς ως το τέλος του χρόνου θα ολοκληρωθεί η μετάβαση του συνόλου της εκτέλεσης των αμερικανικών μελετών στην Αθήνα, γεγονός που θα εκτοξεύσει τον συνολικό αριθμό των απασχολούμενων από 108 τον Απρίλιο σε 150 περίπου έως το τέλος του 2014. Έτσι, σταδιακά το ACE αναδεικνύεται σε αιχμή του δόρατος για τα Analytics της IRI σε παγκόσμιο πλέον επίπεδο. Οι άνθρωποι που αποτύπωσαν στον παγκόσμιο χάρτη των Analytics την Αθήνα και την Ελλάδα είναι στην πλειονότητά τους νέοι επιστήμονες (μ.ο. ηλικίας 32,1 έτη) με σπουδές σε κορυφαία ακαδημαϊκά ιδρύματα. Το 95% είναι κάτοχοι τουλάχιστον ενός μεταπτυχιακού τίτλου σπουδών, ενώ το 18% αυτών κατέχουν

διδακτορικό. Πλέον, πραγματοποιούν 3000+ έργα κάθε χρόνο, σε περισσότερες από 30 χώρες του κόσμου και για περίπου 170 πελάτες παγκοσμίως.

### 5.3 Μετάφραση των δεδομένων σε αξία

Στην εποχή μας όπου η μεγαλύτερη τάση και οι λέξεις που ακούγονται συνεχώς είναι Μεγάλα Δεδομένα (Big Data), τα στελέχη προσπαθούν να καταλάβουν τι μπορεί να σημαίνει αυτό για αυτούς και το μέλλον των επιχειρήσεων τους. Τα δεδομένα δεν έχουν νόημα αν δεν βοηθούν τη λήψη αποφάσεων και αν δεν έχουν ένα μετρήσιμο αποτέλεσμα. Κατά συνέπεια, η μεγάλη πρόκληση είναι η μετατροπή των δεδομένων σε διορατικότητα (insights) και ακόμα περισσότερο σε αξία. Η IRI, καθώς και άλλες συμβουλευτικές εταιρείες που ασχολούνται με τα δεδομένα χρησιμοποιούν διάφορα μοντέλα και προσεγγίσεις για να εξάγουν χρήσιμα συμπεράσματα, προτάσεις για δράση, διορατικότητα (insights) και αξία. Κάποια από αυτά τα μοντέλα, με τα οποία τα στελέχη των επιχειρήσεων μπορούν να χρησιμοποιούν τα δεδομένα για να παίρνουν πιο ακριβείς και αποτελεσματικές αποφάσεις, ακολουθούν παρακάτω μετά τους ορισμούς της αξίας και της διορατικότητας.

#### Αξία

Η αξία είναι το επίτευγμα ενός αποτελέσματος το οποίο είναι σημαντικό, αξίζει ή είναι χρήσιμο σε έναν οργανισμό. Αξία μπορεί να είναι τα έσοδα, το περιθώριο κέρδους, η μείωση κόστους, ο χρόνος, η αποτελεσματικότητα, η απλούστευση, η ακρίβεια, η καινοτομία, η επέκταση, οι δημόσιες σχέσεις, οι επιπτώσεις στο περιβάλλον, η εταιρική ευθύνη. Η αξία έχει διαφορετική σημασία για τον κάθε άνθρωπο που εργάζεται σε έναν οργανισμό ανάλογα με τη θέση που κατέχει ή το τμήμα στο οποίο εργάζεται, για παράδειγμα, για τον εμπορικό διευθυντή μεγαλύτερη σημασία έχουν τα έσοδα, ενώ για τον γενικό διευθυντή μπορεί να έχει μεγαλύτερη σημασία το περιθώριο κέρδους. Ακόμη, μπορεί η ίδια εταιρεία να αλλάζει τον τρόπο που ορίζει την αξία χρόνο με το χρόνο.

## Δεδομένα vs Διορατικότητα (Insights) [5]

Η Merriam-Webster ορίζει τη διορατικότητα (insights) ως τη δύναμη ή τη δράση αυτών που βλέπει κάποιος. Η λέξη κλειδί είναι «βλέπει». Τα δεδομένα θα πρέπει να χρησιμοποιούνται ώστε να εντοπίζονται πλάνα δράσης, τάσεις, ανωμαλίες. Τα δεδομένα δεν βοηθούν να δει κανείς και τα περισσότερα δεδομένα δεν σημαίνουν απαραίτητα καλύτερη διορατικότητα (insights). Το 1990, ο Stephen Tuthill της 3M συνέβαλλε στην σύνδεση των δεδομένων με τη «σοφία». Η «Ιεραρχία των δεδομένων» του εντοπίζει τέσσερα σημαντικά πεδία: τα δεδομένα, τις πληροφορίες, την γνώση και τη σοφία, με τα δεδομένα να είναι οι πρώτες ύλες ή τα γεγονότα. Μόλις υπάρξουν τα δεδομένα, μπορούν να ταξινομηθούν και να οργανωθούν σε πληροφορίες. Η γνώση έρχεται από την ανάλυση, τη σύνθεση και την μετάφραση των μοτίβων και των συσχετίσεων μεταξύ των δεδομένων και άλλων παραγόντων. Η σοφία είναι αποτέλεσμα της κατανόησης του τι έχει σημασία για τον οργανισμό. Έτσι, από την επικέντρωση σε περισσότερα δεδομένα, είναι προτιμότερη η επικέντρωση στον εντοπισμό των σωστών στοιχείων και στη συνέχεια στην ανάλυση τους με τέτοιο τρόπο που θα δώσει τη δυνατότητα για γνώση και για δράση. Παρακάτω περιγράφονται ορισμένες προσεγγίσεις που βοηθούν στη μετατροπή των δεδομένων σε διορατικότητα (insights).

### 5.3.1 Μοντέλο SOLV (Situation, Opportunities, Limitations, Value) – Ανάλυση της κατάστασης, των ευκαιριών, των περιορισμών και της αξίας

Οι περισσότεροι οργανισμοί δεν ζητούν ένα συγκεκριμένο αναλυτικό μοντέλο, συνήθως ζητούν μία λύση για ένα συγκεκριμένο πρόβλημα που αντιμετωπίζουν. Βάσει της IRI, το πιο σημαντικό από όλα, αν θέλει κάποιος του τομέα της ανάλυσης να προσθέσει αξία στον πελάτη του είναι να τον αφήσει να μιλήσει και αυτός να ακούσει καλά το πρόβλημά του. Για να γίνει αυτό θα πρέπει να είναι ανοιχτός στο να δει πως βλέπει η εταιρεία-πελάτης του το πρόβλημα από τη δική της οπτική γωνία, να μην τον διακόψει, να μην αρχίσει να φιλτράρει, να μην τον καθοδηγήσει και προπαντός να μην αρχίσει να απαριθμεί όλες τις λύσεις που προσφέρει η εταιρεία του. Αντίθετα θα πρέπει με τη γλώσσα του σώματος και την οπτική επαφή να δείξει ενδιαφέρον. Με λεκτικά και μη λεκτικά στοιχεία να δημιουργήσει μία συμπάθεια ή μία αρμονική σχέση ώστε να δημιουργήσει ένα κλίμα εμπιστοσύνης. Οι ερωτήσεις που θα γίνουν θα πρέπει να είναι

τέτοιες ώστε να πάρει τη μεγαλύτερη δυνατή πληροφορία, η οποία παράλληλα θα πρέπει να είναι συγκεκριμένη όπως για παράδειγμα «ποια είναι τα τρία βασικά θέματα που σας απασχολούν» και όχι «τι σας απασχολεί» που είναι γενικό. Επιπλέον θα πρέπει να γίνονται ερωτήσεις όπως τι, γιατί, τι θα γίνει αν, τι θα γίνει μετά. Θα πρέπει να γίνεται μία ερώτηση κάθε φορά, αν γίνονται πολλές ερωτήσεις ταυτόχρονα πιθανό κάποιες να μην απαντηθούν ή να μπερδέψουν τον συνομιλητή και κατά συνέπεια τα αιτήματα για την ανάλυση να μην είναι ξεκάθαρα. Ακόμη, για να είναι σίγουρος τι είναι αυτό που του ζητείται, καλό θα είναι να το επιβεβαιώνει κάθε φορά με δικά του λόγια, από τη δική του πλευρά. Αν κάτι δεν είναι κατανοητό θα πρέπει να ζητάει διευκρινήσεις. Όπου χρειάζεται, θα πρέπει να χρησιμοποιεί τη σιωπή, ώστε να αφήνει τον συνομιλητή του να σκεφτεί.

Το δεύτερο βήμα αφού ακούσει και συζητήσει τα θέματα του πελάτη του, θα πρέπει να σκιαγραφήσει το πλαίσιο της παρούσας κατάστασής του. Για παράδειγμα, ποιο είναι το αντικείμενο του πελάτη του, ποια είναι η παρούσα θέση του σε σχέση με τον τομέα στον οποίο δραστηριοποιείται και σε σχέση με τα θέματα που έχουν τεθεί υπό συζήτηση, ποιο είναι το στρατηγικό του περιβάλλον και πως η εταιρεία ανταποκρίνεται σε αυτό. Ποιοι είναι οι πελάτες του, υπάρχει κάποια αλλαγή της συμπεριφοράς τους το τελευταίο διάστημα, ποιες είναι οι εξωτερικές τάσεις που μπορεί να τον επηρεάσουν. Σε αυτό το στάδιο μπορεί να πάρει αρκετές πληροφορίες από τον πελάτη όπως ποιοι είναι οι βασικοί δείκτες αποτελεσματικότητας που χρησιμοποιούνται για να μετρηθεί η αποτελεσματικότητα της εταιρείας, τι περιμένουν οι βασικοί πελάτες από αυτούς, ποιοι είναι οι βραχυπρόθεσμοι και ποιοι οι μεσοπρόθεσμοι στόχοι τους, τι πιέσεις δέχονται από τον ανταγωνισμό, ποια θεωρούν δυνατά τους σημεία και ποια αδύναμα, σε ποιες καινοτομίες επικεντρώνονται. Ακόμη, πως οι αλλαγές του οικονομικού περιβάλλοντος επηρεάζουν αυτούς και τους πελάτες τους. Τι χρειάζεται η συγκεκριμένη εταιρεία να κάνει καλύτερα.

Το τρίτο βήμα αφορά την διερεύνηση των ευκαιριών, την εύρεσή τους και τον καθορισμό της αξίας που θα έχουν για το άτομο και την εταιρεία από την οποία έχει κληθεί να λύσει το πρόβλημα. Σε αυτή τη φάση, θα πρέπει να διερευνήσει επίσης ποιοι είναι οι τωρινοί στόχοι, ποιοι είναι οι μελλοντικοί στόχοι, ποια θα πρέπει να είναι τα αποτελέσματα, ποιες οι επιπτώσεις. Ποια μπορεί να είναι η οικονομική ευκαιρία και ποιοι οι μοχλοί που θα συμβάλουν στην εκμετάλλευσή της. Θα πρέπει να μάθει από τις προκλήσεις που έχουν συζητηθεί ποιες είναι οι πιο σημαντικές και γιατί. Από τις ευκαιρίες που έχουν συζητηθεί ποιες είναι οι πιο σημαντικές. Τι σημαίνει η εκμετάλλευσή αυτών των ευκαιριών. Ποιο αποτέλεσμα θα είχε μεγαλύτερη σημασία αν

οι προκλήσεις που συζητήθηκαν θα μπορούσαν να ξεπεραστούν. Τι περιμένουν να δουν βραχυπρόθεσμα και τι μακροπρόθεσμα, αν οι προκλήσεις αυτές ξεπεραστούν. Ποια άλλα παρεπόμενα οφέλη θα έχει ο οργανισμός αν λυθούν τα θέματα που συζητήθηκαν. Από όλα τα παραπάνω θα καταφέρει να ξεχωρίσει τις πιο υποσχόμενες ευκαιρίες για τον πελάτη, την αξία που συνδέεται με την κάθε ευκαιρία και να διερευνήσει αν ο πελάτης τις πιστεύει.

Το τέταρτο βήμα έχει να κάνει με τους περιορισμούς, τι θα εμποδίσει το άτομο και τον οργανισμό του από το να πραγματοποιήσει τις ευκαιρίες ή να πετύχει τα αναμενόμενα αποτελέσματα. Πιθανά εμπόδια μπορεί να είναι οι περιορισμένοι πόροι, ο προϋπολογισμός, η κουλτούρα. Για να μάθει κάποιος τους περιορισμούς θα πρέπει να ρωτήσει τι ανησυχεί τον πελάτη του περισσότερο, τι μπορεί να εμποδίσει την επιτυχημένη εφαρμογή της συγκεκριμένης λύσης, πως θα μπορούσε να ξεπεράσει τα θέματα που μπορεί να τον εμποδίσουν, πως νιώθει στο να κάνει αλλαγές, το συγκεκριμένο θέμα που τον απασχολεί έχει ξανασυζητηθεί στο παρελθόν, γιατί όχι, ή αν έχει συζητηθεί τι διαφορετικό συμβαίνει αυτή τη φορά. Πόσο δύσκολο είναι να βρεθούν οι πόροι που μπορεί να χρειαστούν για το συγκεκριμένο έργο. Σε αυτό το στάδιο φαίνονται οι πόροι που μπορεί να χρειαστούν για να εκμεταλλευτούν οι ευκαιρίες που εντοπίστηκαν. Μπορεί να δει ποιες ευκαιρίες μπορεί να εκμεταλλευτεί ο πελάτης μόνος του και σε ποιες χρειάζεται τη βοήθεια ενός εξωτερικού συνεργάτη, καθώς και αν οι περιορισμοί μπορούν να ξεπεραστούν.

Τέλος, προτείνεται μία προσέγγιση με την οποία μπορεί να υλοποιηθεί η ευκαιρία που έχει εντοπιστεί και συμφωνούνται οι πιθανοί μέθοδοι για τον καθορισμό της οικονομικής αξίας του επιθυμητού αποτελέσματος. Ακόμη και σε αυτό το στάδιο μπορεί να τον ρωτήσει ποια θεωρεί τα τρία πιο σημαντικά στάδια που θα ήθελε να είναι κομμάτι της εφαρμογής του συγκεκριμένου έργου.

Με όλα τα παραπάνω καταφέρνει να δει ποιο είναι το πραγματικό πρόβλημα που απασχολεί τον πελάτη και να του προτείνει μία στοχευμένη λύση που θα έχει αξία για αυτόν. Παράλληλα, θα έχει σκιαγραφήσει το πλαίσιο, τις πιο σημαντικές ευκαιρίες και τα οφέλη, καθώς και τους περιορισμούς που μπορεί να προκύψουν. Δεν έχει νόημα να προτείνει μία λύση ή να σπαταληθούν πόροι για ένα αναλυτικό μοντέλο και οι δράσεις να μην μπορούν να υλοποιηθούν επειδή δεν έχουν εξεταστεί τα πιθανά εμπόδια. Έτσι η λύση δεν θα είχε καμία αξία για τον πελάτη.

### 5.3.2 Ανάλυση SWOT (Strengths, Weaknesses, Opportunities, Threats)

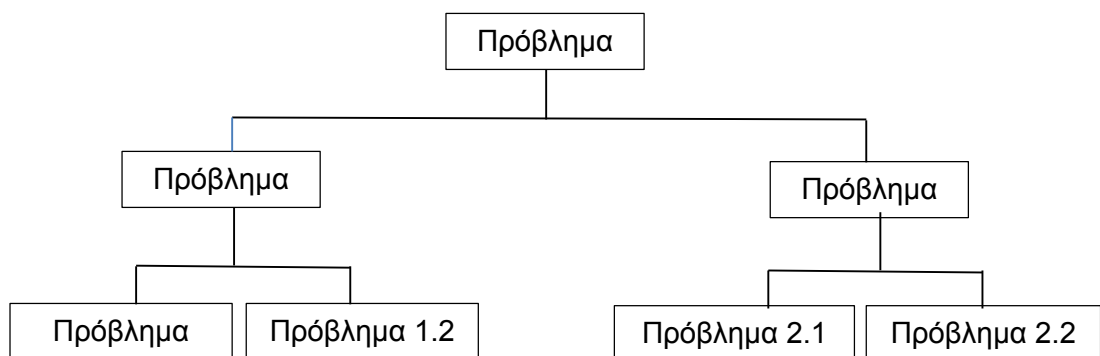
Προχωρώντας στην παρουσίαση και ερμηνεία των αποτελεσμάτων, η IRI χρησιμοποιεί τη γνωστή ανάλυση SWOT με τον τρόπο που περιγράφεται στη συνέχεια. Περιγράφεται η κατάσταση του πελάτη, η παρούσα δομή του, η θέση του στην αγορά, το περιβάλλον του, οι βασικές προοπτικές, οι στρατηγικές, οι στόχοι ανάπτυξης και οι κίνδυνοι. Αναφέρονται οι σημαντικότερες δυνάμεις, αδυναμίες, ευκαιρίες και απειλές σε σχέση με ένα συγκεκριμένο θέμα, τμήμα ή οργανισμό. Στη συνέχεια διατυπώνονται τα βασικά συμπεράσματα.

Με βάση την ανάλυση SWOT, αναζητούνται αρχικά οι απαραίτητες πληροφορίες, αναλύσεις, εφαρμογές σε όλα τα τμήματα, λειτουργίες ή δραστηριότητες ενός οργανισμού. Σκιαγραφώντας τις δυνάμεις, δημιουργείται μία λίστα με τα χαρακτηριστικά της εταιρείας που θεωρούνται κεφάλαια (assets) για την επιχείρηση. Αυτά θα μπορούσαν να είναι η μάρκα, η θέση στην αγορά, τα κανάλια διανομής, μία καινοτομία, ο οργανισμός, οι λειτουργίες. Από την άλλη πλευρά τα παραπάνω στοιχεία θα μπορούσαν να αποτελούν αδυναμίες. Σχετικά με τις ευκαιρίες, θα πρέπει να αποτυπωθούν τα σημεία που βλέπει ο εξωτερικός συνεργάτης ότι θα μπορούσαν να αποτελέσουν δυναμικές ανάπτυξης και τα αντίστοιχα σημεία από την οπτική γωνία του πελάτη. Στις απειλές, διατυπώνονται εκείνα τα σημεία που αποτελούν εμπόδιο στην επιτυχία μίας δράσης ή μίας στρατηγικής, ποια σημεία βλέπει ο εξωτερικός συνεργάτης και ποια ο πελάτης.

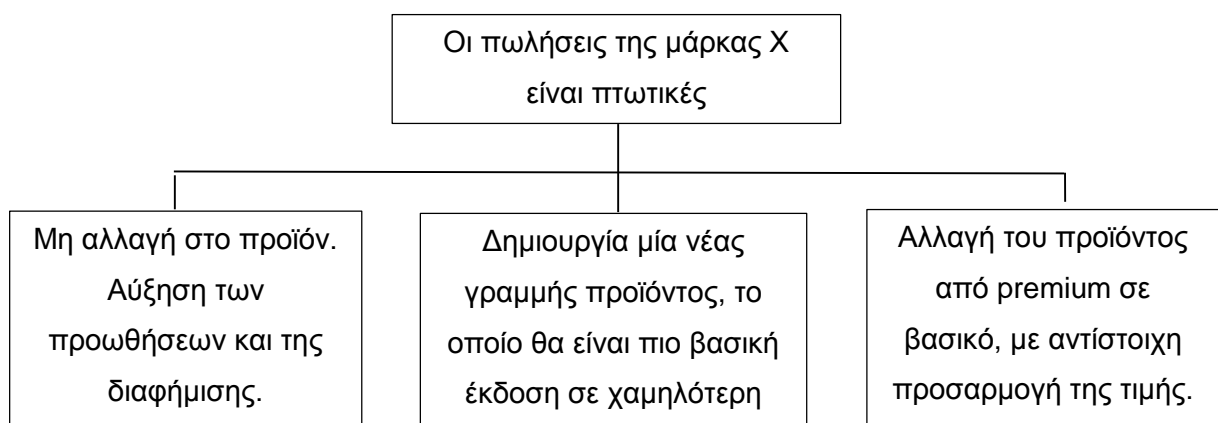
Παράλληλα, είναι σημαντικό να εξεταστεί το ανταγωνιστικό περιβάλλον. Εξετάζονται οι τρεις βασικοί ανταγωνιστές, συμπεριλαμβανομένων τόσο των παραδοσιακών όσο και των μη παραδοσιακών και το αντιλαμβανόμενο ή πραγματικό ανταγωνιστικό τους πλεονέκτημα. Δημιουργείται μία λίστα με τις δυνάμεις και τις αδυναμίες του κάθε ανταγωνιστή. Στη συνέχεια εξετάζεται ποιοι είναι οι βασικοί αγοραστές του πελάτη, ποιοι χρησιμοποιούν τα προϊόντα του και γιατί, τι κάνει αυτά τα προϊόντα μοναδικά, πως ο πελάτης διαφοροποιείται στην αγορά. Ποια είναι τα κανάλια διανομής του και ποιοι είναι οι λιανέμποροι, ποια η σχέση του μαζί τους. Άλλα θέματα που χρειάζεται να διερευνηθούν είναι ο καθορισμός των προτεραιοτήτων για περικοπές κόστους, ανάπτυξη, καινοτομία, απάντηση στον ανταγωνισμό.

Αφού καθοριστούν τα παραπάνω, τότε ακολουθεί ο σχηματισμός υποθέσεων για να μπει το πρόβλημα σε ένα πλαίσιο. Οι υποθέσεις που χρησιμοποιούνται προέρχονται

από την πλευρά της αγοράς, του καταναλωτή και της ανάλυσης. Αρχικά, γίνεται επανεξέταση του προβλήματος, το οποίο έχει επιλεγεί και εύρεση των πιθανών λόγων στους οποίους μπορεί να οφείλεται. Οι λόγοι μπορεί να προκύψουν από τη δύναμη ή αδυναμία της μάρκας, από τις αλλαγές στην καταναλωτική συμπεριφορά, από το κοινωνικοοικονομικό περιβάλλον, από την επικοινωνία, από τις προσφορές. Στη συνέχεια, γίνεται διαχωρισμός του προβλήματος σε διαχειρίσιμα και συγκεκριμένα κομμάτια. Διευκρινίζονται οι προτεραιότητες ώστε να γίνει επικέντρωση σε συγκεκριμένους στόχους. Γίνεται δηλαδή επιλογή ενός κομματιού, μίας ομάδας υποθεμάτων και από αυτή ξεχωρίζεται το κύριο υπό-θέμα. Γίνεται αναζήτηση και ταξινόμηση των λόγων που μπορεί να ευθύνονται για αυτό το υπό – θέμα. Στη συνέχεια, γίνεται ανεύρεση πιθανών στοιχείων για να επιβεβαιωθούν οι υποθέσεις και τα θέματα στα οποία έχει καταλήξει ο αναλυτής. Η παραπάνω διαδικασία επιτρέπει την καλύτερη κατανόηση του προβλήματος, καθώς αν χρειάζονται επιπλέον πληροφορίες, στοιχεία ή πόροι για να προταθεί ένα σχέδιο δράσης. Επιπλέον, εντοπίζονται οι λόγοι που οδηγούν στο συγκεκριμένο πρόβλημα και οι συνέπειές του. Η παραπάνω διαδικασία, απεικονίζεται με τη χρήση του «Δέντρου Θεμάτων» :



Είναι πολύ σημαντικό να διαμορφωθεί η αρχική υπόθεση και να υπάρχει η δυνατότητα μετατροπής των υποθεμάτων σε υποθέσεις που να μπορούν να μετρηθούν. Κάποιες από αυτές τις υποθέσεις μπορεί να είναι λάθος και να χρειάζεται να επανεξεταστούν ή να διερευνηθούν περαιτέρω. Παρακάτω είναι ένα παράδειγμα όπου υπάρχει το πρόβλημα και η δημιουργία τριών υποθέσεων:



Το επόμενο βήμα είναι η διαμόρφωση των ευρημάτων. Η υπόθεση που θα επιλεγεί ως η πιο σημαντική είναι κι αυτή που θα καθορίσει την ανάλυση. Στην ανάλυση θα πρέπει να βρεθούν οι βασικοί μοχλοί για την υλοποίηση της υπόθεσης. Δεν χρειάζεται να χαθεί κανείς στον ωκεανό των δεδομένων, απλά θα πρέπει να επικεντρωθεί γύρω από τα βασικά θέματα. Για να φτιαχτεί ένα πλάνο δράσης θα πρέπει να χρησιμοποιηθούν πληροφορίες από διάφορες πηγές.

Πέρα από τα δεδομένα που θα χρησιμοποιηθούν σε ένα αναλυτικό μοντέλο, θα πρέπει να ληφθούν υπόψη και πληροφορίες από άλλες πηγές. Η εξέταση δεδομένων από διαφορετικές πηγές είναι μία πάρα πολύ βασική διαδικασία, αφού επιτρέπει μία ολιστική εικόνα της κατάστασης. Το στέλεχος που καλείται να προτείνει ένα πλάνο δράσης το οποίο θα έχει νόημα και θα προσφέρει αξία στον πελάτη του, θα βοηθηθεί πολύ αν μιλήσει ο ίδιος με κάποιους από του πελάτες του πελάτη του, ώστε να καταλάβει τη δική τους οπτική γωνία σχετικά με το προϊόν που είναι υπό εξέταση. Μπορεί να επισκεφτεί τα κανάλια διάθεσης των προϊόντων του πελάτη του και να δει από κοντά πως είναι τοποθετημένα, ποιοι είναι οι τύποι προώθησης που χρησιμοποιούνται, ποιο είναι το κωδικολόγιο της μάρκας που είναι τοποθετημένο. Μπορεί για παράδειγμα, να γίνεται ένα αναλυτικό μοντέλο ώστε να εξεταστεί η αποτελεσματικότητα των προωθητικών ενεργειών μέσα στα σημεία πώλησης και το δεύτερο σημείο πώλησης μέσα στο κατάστημα, ενώ γίνεται σε μεγάλη κλίμακα καταστημάτων, να φαίνεται ότι δεν είναι αποδοτική ενέργεια. Αν το στέλεχος έχει επισκεφθεί κάποια από τα καταστήματα και δει ότι το δεύτερο σημείο πώλησης βρίσκεται σε σημείο όπου δεν υπάρχει μεγάλη διέλευση, είναι λογικό να μην βγαίνει αυτή η ενέργεια αποδοτική, παρόλο που γίνεται σε μεγάλη κλίμακα καταστημάτων. Αυτό είναι ένα ποιοτικό χαρακτηριστικό και μπορεί να το χρησιμοποιήσει κατά την ερμηνεία των αποτελεσμάτων του αναλυτικού μοντέλου. Άλλες πηγές που μπορεί να χρησιμοποιήσει είναι το διαδίκτυο ή το περιβάλλον. Όλα τα παραπάνω θα τον βοηθήσουν να κάνει μία καλή κριτική των αποτελεσμάτων, τα οποία προέρχονται από τα αναλυτικά μοντέλα και αποτελούν τη λογική ανάλυση των δεδομένων, η οποία έχει προέλθει από την εφαρμογή κάποιων αλγορίθμων ή μαθηματικών.

Το επόμενο στάδιο, μετά την αποτύπωση και ερμηνεία των ευρημάτων, είναι η δημιουργία διορατικών λύσεων (insights). Οι λύσεις αυτές υποστηρίζουν μία συγκεκριμένη απόφαση ή συμβάλλουν στην καλύτερη κατανόηση μίας συγκεκριμένης κατάστασης και είναι επωφελείς για τις επιχειρήσεις. Διορατικότητα (insight) είναι μία σκέψη, ένα γεγονός, ένας συνδυασμός γεγονότων, στοιχεία ή/και ανάλυση στοιχείων:



1. Που αλλάζει τα δεδομένα, αλλάζει τον τρόπο σκέψης, οδηγεί κάπου αλλού.
2. Συμβάλλει στην κατανόηση ή στην περαιτέρω κατανόηση μίας κατάστασης ή ενός θέματος.
3. Ωφελεί την επιχείρηση ή αλλάζει τον τρόπο σκέψης για μία κατάσταση ή για ένα συγκεκριμένο θέμα.
4. Αλλάζει τη σκέψη σχετικά με τη δυναμική της επιχείρησης και συνήθως αλλάζει τις αποφάσεις.

Μπορεί επίσης να είναι και μία δήλωση που προέρχεται από ένα βαθύτερο νόημα, το οποίο είναι απόλυτα κατανοητό στο κοινό-στόχο μία μάρκας. Μπορεί να έχει να κάνει με τη στάση, τις πεποιθήσεις και να συνδέεται συναισθηματικά με τον καταναλωτή, προκαλώντας μία άμεση απάντηση/αντίδραση.

Αφού ο αναλυτής καταλήξει στις διορατικές λύσεις, θα πρέπει να κάνει και κριτική πάνω σε αυτές. Θα τις καταλάβουν οι αποδέκτες, είναι κάτι καινούργιο, είναι κάτι αρκετά απλό, μπορεί να το υποστηρίξει; Το τελευταίο στάδιο είναι οι προτάσεις για δράση που βοηθούν τους πελάτες να βασιστούν πάνω στις διορατικές λύσεις (insights) και να δράσουν.

#### Παράδειγμα ευρήματος – διορατικότητας (insight) – πρότασης για δράση

##### Στόχος

Ανάπτυξη των πωλήσεων της μάρκας X σε όγκο χωρίς να υπονομεύεται η κερδοφορία μακροπρόθεσμα.

##### Ευρήματα

Αυτό που βλέπει κάποιος από τα δεδομένα είναι ότι (τι έχει συμβεί):

- Η κατηγορία είναι σταθερή σε όγκο, αλλά η μάρκα X αναπτύσσεται 4%. Η ανάπτυξη αυτή προέρχεται από προωθητικές ενέργειες.
- Υπάρχουν κενά διανομής σε συγκεκριμένους τύπους καταστημάτων και κανάλια.
- Οι προωθητικές ενέργειες μειώνουν τα περιθώρια κέρδους κατά 33%.

##### Διορατικότητα (insights)

Πόσο σημαντικά είναι τα ευρήματα, πως επηρεάζουν το αίτημα που δόθηκε προς διερεύνηση (γιατί έχουν σημασία, πως επηρεάζουν τη στρατηγική ή τις τακτικές):

- Ενώ οι πωλήσεις σε όγκο της μάρκας αναπτύσσονται με ρυθμό 4% λόγω των προωθητικών ενεργειών, η ανάλυση της κερδοφορίας δείχνει ότι αυτή δεν είναι βιώσιμη μακροπρόθεσμα, καθώς οι καταναλωτές θα συνηθίσουν στις

προσφορές και θα τις περιμένουν. Δεν θα ψωνίζουν αν η μάρκα δεν είναι σε προώθηση. Αυτό θα έχει σαν αποτέλεσμα την κατάρρευση της κερδοφορίας των βασικών πωλήσεων.

Προτάσεις για δράση

Ευκαιρίες που αναφέρονται στις διορατικές λύσεις (insights), (τι πρέπει να κάνουμε για αυτό):

- Εξέταση της προοπτικής να αναπτυχθεί η διανομή και σε άλλα κανάλια, όπου θα πωλούνται τα προϊόντα με το κανονικό περιθώριο κέρδους, ώστε να «χτιστεί» από εκεί η κερδοφορία των βασικών πωλήσεων

### 5.3.3 Προσέγγιση «Insights Factory» από τη McKinsey [4]

Σύμφωνα με τη McKinsey, τα Μεγάλα Δεδομένα (Big Data) δεν φτάνουν στα στελέχη της πρώτης γραμμής σε σχετικές και χρήσιμες μορφές. Πολλά έργα με Μεγάλα Δεδομένα (Big Data) δημιουργούνται χωρίς την εμπλοκή των στελεχών της πρώτης γραμμής ή χρειάζονται πάρα πολύ χρόνο για να ολοκληρωθούν που ίσως μερικές φορές να είναι αργά, για τις αποφάσεις που πρέπει να ληφθούν. Έτσι, για τη δημιουργία αξίας από τα Μεγάλα Δεδομένα (Big Data) προτείνεται να συνδέονται τα δεδομένα με διορατικές λύσεις (insights) και προτάσεις για δράση γρήγορα και επαναλαμβανόμενα. Η McKinsey προτείνει την προσέγγιση του εργοστασίου, όπου οι διορατικές λύσεις (insights) είναι το τελικό προϊόν, τα δεδομένα είναι η πρώτη ύλη και τα στελέχη της πρώτης γραμμής είναι οι καταναλωτές. Η συγκεκριμένη προσέγγιση δίνει τη δυνατότητα στις εταιρείες να φιλτράρουν τον τεράστιο όγκο των δεδομένων γρήγορα, να εφαρμόζουν το κατάλληλο αναλυτικό μοντέλο και να παρέχουν σχετικές λύσεις, ώστε τα στελέχη να παίρνουν τις σωστές αποφάσεις. Η συγκεκριμένη προσέγγιση περιλαμβάνει τα παρακάτω τέσσερα στάδια:

#### 1.Απόφαση του τι θα παραχθεί

Πριν αρχίσει κάποιος να εργάζεται πάνω σε ένα έργο, θα πρέπει να έχει μία ξεκάθαρη εικόνα του τι θέλει να πετύχει. Θα πρέπει να έχει αποφασίσει ποιες είναι οι συγκεκριμένες ερωτήσεις που η επιχείρησή του χρειάζεται να απαντήσει για ένα θέμα, και οι απαντήσεις που θα δοθούν ποιες δράσεις θέλει να πυροδοτήσουν. Να βάλει σε προτεραιότητα ερωτήσεις που έχουν να κάνουν με τις μεγαλύτερες οικονομικές ευκαιρίες και θα οδηγήσουν σε εφαρμόσιμες δράσεις. Στη συνέχεια θα πρέπει να

ρυθμίσει το «εργοστάσιό» του να παραγάγει αυτές ακριβώς τις διορατικές λύσεις (insights). Για παράδειγμα, ένας λιανέμπορος ανακάλυψε ότι το 90% της πτώσης των πωλήσεων του χρόνο με τον χρόνο προέρχονταν από ένα 12% των πελατών σε συγκεκριμένες αγορές. Έτσι, οι ερωτήσεις επικεντρώθηκαν, αντιστοίχως, στην κατανόηση των λόγων που συνέβαινε αυτό και ο λιανέμπορος κατάφερε γρήγορα να αντιστρέψει την τάση με στοχευμένες τακτικές merchandising στις συγκεκριμένες αγορές.

## 2. Άντληση των πρώτων υλών

Αν και θα ήταν χρήσιμο να βρει κανείς ένα εύρος πηγών που θα συμβάλλουν στη δημιουργία των διορατικών λύσεων (insights), καλό θα είναι να ξεκινήσει άμεσα με τα καλύτερα διαθέσιμα δεδομένα. Η αναζήτηση των τέλειων δεδομένων είναι μία χρονοβόρα διαδικασία, συχνά δεν αποδίδει, και μειώνει την δυνατότητα γρήγορης δράσης. Η McKinsey προτείνει να ξεκινήσει το έργο με μία μικρή ομάδα στοιχείων. Μία ολοκληρωμένη «αποθήκη στοιχείων» είναι μεγάλο κεφάλαιο μακροπρόθεσμα για την επιχείρηση αλλά με μία μικρότερη, πιο επιλεκτική ομάδα στοιχείων είναι πιο εύκολο να παραχθούν οι διορατικές λύσεις (insights) πιο γρήγορα, αποτρέποντας το χάσιμο χρόνου μέσα στην πολυπλοκότητα. Με τον καιρό κάποιος μπορεί να προσθέτει κι άλλες ομάδες δεδομένων. Για παράδειγμα, ένας κορυφαίος λιανέμπορος, θέλοντας να καταλάβει την αγοραστική συμπεριφορά των πελατών του χρησιμοποίησε τα δικά του στοιχεία συναλλαγών από τα σημεία του, στοιχεία από τρίτους που τα ομαδοποιούσαν, κοινά δεδομένα από ανταγωνιστές και δημόσιες πηγές. Όλα αυτά τα στοιχεία ήταν άμεσα διαθέσιμα. Σε ένα χρόνο εμπλούτισε τη διορατικότητά του προσθέτοντας στοιχεία από τα μέσα κοινωνικής δικτύωσης (για ανάλυση συναισθήματος), στοιχεία τοποθεσίας (για να καταλάβει την κίνηση στο κατάστημα) και οικονομικές πληροφορίες από παρόχους πιστωτικών καρτών (για το μερίδιο πορτοφολιού).

## 3. Παραγωγή διορατικών λύσεων (insights) με ταχύτητα

Σύμφωνα με τη McKinsey, όταν φτάνουμε στο κομμάτι του αναλυτικού μοντέλου, η παραγωγική διαδικασία είναι προϊόν ταχύτητας. Επικέντρωση σε γρήγορες αποφάσεις και εκτέλεση, που δεν επιτρέπουν μακροσκελείς συζητήσεις, και οδηγούν σε διορατικές λύσεις (insights) που τα στελέχη της πρώτης γραμμής μπορούν να χρησιμοποιήσουν. Πρέπει να μπουν πεπερασμένα χρονικά όρια στο «εργοστάσιο insights», ώστε σε σύντομους χρόνους παραγωγής να δημιουργηθεί ένα δομημένο αποτέλεσμα βάσει των επαναλαμβανόμενων αναλυτικών μοντέλων και του αυτοματισμού. Συστήνει τη λειτουργία σαν start up που έχει ανάγκη από ταχύτητα και έτσι δεν περιμένει για το τέλειο, της αρκεί το αρκετά καλό. Επιπλέον, είναι καλό να

δημιουργούνται μικρές, ευέλικτες ομάδες που συνδυάζουν στρατηγικές, αναλυτικές και τεχνικές ικανότητες για συγκεκριμένες θεματικές ενότητες και όχι μια ενιαία, γενικευμένη, και συνήθως αργοκίνητη ομάδα. Για να δουλεύει το εργοστάσιο όλο το εικοσιτετράωρο, συστήνεται η πρόσληψη εξωτερικού συνεργάτη που θα εκτελεί δομημένη ανάλυση συνεχώς, με σχετικά χαμηλό κόστος.

#### 4. Παράδοση των «προϊόντων» και δράση

Η «αρκετά καλή πληροφόρηση» είναι διαθέσιμη άμεσα και μπορεί να χρησιμοποιηθεί άμεσα για συγκεκριμένες δράσεις. Αν από τα στοιχεία βγει το συμπέρασμα ότι υπάρχει 90% πιθανότητα το γάλα και τα αυγά να αγοράζονται μαζί, γιατί να μην εφαρμοστεί γρήγορα μία πιλοτική τοποθέτηση του γάλακτος δίπλα στα αυγά από το να αναζητώνται πιο ολοκληρωμένες επιλογές.

Για να οδηγήσουν οι διορατικές λύσεις (insights) σε δράσεις, πρέπει να υπάρχει μία ξεκάθαρη εικόνα του τι τα στελέχη της πρώτης γραμμής μπορούν να χρησιμοποιήσουν στην πραγματικότητα. Αυτά τα στελέχη θα πρέπει να βρουν τι χρειάζονται. Συχνά, οι μαρκετίερς ή τα στελέχη πωλήσεων παίρνουν στοιχεία που στο τέλος τα αγνοούν. Σε πολλές περιπτώσεις η ανάλυση δεν είναι πρακτική, δεν είναι ξεκάθαρη, δεν είναι αξιόπιστη ή δεν θεωρείται σχετική. Για ένα «εργοστάσιο insights» αυτά τα στελέχη, είναι οι πελάτες. Θα πρέπει να παίρνουν μέρος στην παραγωγική διαδικασία και να χρησιμοποιούν τη διορατικότητα με απλούς τρόπους, όπως με τα διαδραστικά εργαλεία. Δεν χρειάζεται να πιεστούν για να τα χρησιμοποιήσουν, απλά θα πρέπει ο αναλυτής να τους ακούσει και να ανταποκριθεί στις ανάγκες τους και έτσι θα τα ζητήσουν από μόνοι τους.

#### Δημιουργία ενός «εργοστασίου» κουλτούρας διαχρονικά

Για να λειτουργήσει το εργοστάσιο διορατικών λύσεων (insights) με τον τρόπο που οι επιχειρήσεις δουλεύουν, δεν θα πρέπει να γίνονται μεγάλες αλλαγές. Θα πρέπει να συνηθίσουν τα ενδιαφερόμενα μέρη να χρησιμοποιούν σταδιακά και προοδευτικά τα δεδομένα και τις διορατικές λύσεις (insights) στην καθημερινή τους διαδικασία λήψης αποφάσεων. Με τον χρόνο, όταν η διορατικότητα γίνει κάτι συνηθισμένο, τα στελέχη θα είναι πιο πρόθυμα να πάρουν μεγαλύτερες αποφάσεις και να κάνουν μεγαλύτερες αλλαγές.

### 5.3.4 Πέντε τρόποι για την μετατροπή των Μεγάλων Δεδομένων σε διορατικότητα και δράση [5]

Σύμφωνα με τη Laura Patterson, και το άρθρο της στο online marketing institute, προτείνονται τα πέντε παρακάτω στάδια για την εξαγωγή διορατικών λύσεων (insights) μέσω των δεδομένων:

#### 1. Δίνουμε στα δεδομένα ζωή μέσω της οπτικοποίησης

Το να έχει κανείς δεδομένα είναι ένα πράγμα, το να τα αναλύει και να τα συνθέτει είναι ένα άλλο. Η σύνθεση αποτελεί το στάδιο όπου φαίνονται τα πρώτα σχέδια δράσης. Μόλις η σύνθεση ολοκληρωθεί, χρειάζεται ένας τρόπος για να δοθεί ζωή στα δεδομένα. Η οπτικοποίηση των δεδομένων είναι κάτι που βοηθάει πάρα πολύ σε αυτό το στάδιο της διαδικασίας. Η οπτικοποίηση παρουσιάζει τα αποτελέσματα της ανάλυσης και δίνει τη δυνατότητα να εντοπιστεί πιο εύκολα η σχέση μεταξύ των μεταβλητών, να εντοπιστούν και να επικοινωνηθούν σημαντικά πλάνα δράσης, ακόμα και να υποστηριχτούν μοντέλα προβλέψεων. Η οπτικοποίηση είναι ένα σημαντικό στάδιο για τον εντοπισμό τάσεων και πλάνων δράσης που διαφορετικά μπορεί να μην φαίνονταν.

#### 2. Συζήτηση των σχεδίων δράσης και των πιθανών συμπερασμάτων

Όλα τα σχέδια δράσης δεν είναι σχετικά. Χρειάζεται χρόνος για να επανεξεταστεί και να συζητηθεί το κάθε πλάνο δράσης και τα πιθανά επακόλουθά του. Πρέπει να εξεταστεί γιατί το κάθε πλάνο δράσης είναι σημαντικό και τι σημαίνει. Αυτό είναι ένα σημαντικό σκαλοπάτι για τη μετάβαση από τις πληροφορίες στη γνώση.

#### 3. Διατύπωση των διορατικών λύσεων (insights) που προκύπτουν από το κάθε πλάνο δράσης

Με μία απλή πρόταση θα πρέπει να διατυπωθεί η διορατική λύση (insight) που έχει προκύψει από το κάθε σχέδιο δράσης ή κομμάτι της σύνθεσης. Μία λύση μπορεί να είναι να γράφονται οι ιδέες σε ένα post-it και να κολλιέται στον τοίχο ή σε ένα flip chart, ώστε εύκολα να ακολουθείται η κάθε λύση και να φαίνεται η μεγάλη εικόνα που συντίθεται καθώς προχωράει το έργο.

#### 4. Δίνουμε χρόνο στα insights να «εκκολαφθούν»

Χρειάζεται να παρθούν αποστάσεις από τις διορατικές λύσεις (insights), να δοθεί τουλάχιστον μία μέρα χρόνος στον αναλυτή και στην ομάδα του. Όταν τα ξαναδούν με φρέσκο μυαλό, θα αποφασίσουν αν χρειάζεται να κάνουν αλλαγές ή όχι.

#### 5. Ρωτάμε και τους άλλους

Έχουν νόημα οι διορατικές λύσεις (insights), οι άλλοι τις καταλαβαίνουν, συμφωνούν; Μόλις θα υπάρχει άνεση με τα συμπεράσματα και τις λύσεις στις οποίες κατέληξε ο αναλυτής θα πρέπει να τις συζητήσει και με τους άλλους ανθρώπους που είχαν λάβει μέρος στα αρχικά στάδια του έργου ώστε να δει τις αντιδράσεις τους, αφού τους ξαναβάλει στο πλαίσιο του έργου. Σε αυτό το στάδιο μπορεί να αποφασίσει αν οι διορατικές λύσεις (insights) στις οποίες κατέληξε έχουν νόημα και είναι αρκετές για να επηρεάσουν ή να οδηγήσουν σε σημαντικές αποφάσεις.

Η επιτυχία αυτής της προσέγγισης έχει να κάνει με την ποιότητα και όχι απαραίτητα με την ποσότητα των δεδομένων.

#### 5.3.5 Από τα δεδομένα σε διορατικές λύσεις (insights): Το Blueprint για μία επιχείρηση [6]

Ο Daniel Waisberg της Google, προτείνει τα παρακάτω στάδια που συμβάλλουν στην μετάφραση των δεδομένων σε διορατικές λύσεις (insights):

##### 1. Καθορισμός των στοιχείων

Για να καταλήξει ο αναλυτής σε επιτυχημένες λύσεις, θα πρέπει να ξέρει τι θέλει από τα δεδομένα - να έχει βρει την αξία τους. Ένα έργο είναι όπως μία ιστορία, έχει το πλαίσιο, την ανάγκη, το όραμα και ένα αποτέλεσμα. Οι παρακάτω ερωτήσεις συμβάλλουν στο καλύτερο πλάνο για τη χρήση των στοιχείων:

Πλαίσιο: Τι προσπαθεί να πετύχει ο αναλυτής; Ποιος επενδύει στα αποτελέσματα του έργου; Υπάρχουν κάποιοι μεγαλύτεροι στόχοι ή προθεσμίες που μπορεί να συμβάλλουν στο να μπει σε προτεραιότητα το έργο;

Ανάγκη: Ποιες ανάγκες μπορούν να καλυφθούν με την έξυπνη χρήση των στοιχείων; Τι θα επιτευχθεί με αυτό το έργο που πριν δεν ήταν εφικτό;

Όραμα: Μπορεί να γίνει ένα ομοίωμα του τελικού αποτελέσματος; Ποια είναι η λογική πίσω από τη λύση;

Αποτέλεσμα: Πως και από ποιον θα χρησιμοποιηθεί το αποτέλεσμα και πως θα ενσωματωθεί μέσα στην εταιρεία;

## 2. Δημιουργία του πλαισίου

Σε αυτό το στάδιο θα πρέπει να βρεθεί η ιστορία που θα πει το έργο. Οι αρχιτέκτονες είναι ένα καλό παράδειγμα. Όπως από αυτούς ζητείται να γίνει ένα αφηρημένο σχέδιο πραγματικότητα, έτσι και οι μαρκετίερς θα πρέπει να μετατρέψουν τα δεδομένα σε λύσεις (insights). Η κατανόηση και η οπτικοποίηση των δεδομένων, είναι όπως όταν ένας αρχιτέκτονας φτιάχνει τα σκίτσα στο αρχικό στάδιο. Η δημιουργία πολλαπλών γραφημάτων και πινάκων, είναι σαν να υπάρχει ένα είδος επικοινωνίας με τα στοιχεία, συμβάλλει στη γνωριμία με τα νούμερα ώστε να βρεθούν ενδιαφέρουσες διορατικές λύσεις (insights).

Σε ένα άρθρο των New York Times, ο Michael Graves, καθηγητής αρχιτεκτονικής στο Princeton, αποκρούει την ιδέα ότι η τεχνολογία προκάλεσε «τον θάνατο της ζωγραφικής» στην αρχιτεκτονική. Επισημαίνει ότι η ζωγραφική είναι μία διαδικασία όπως ένα τελικό προϊόν, και όπως οι υπολογιστές έχουν μία θέση, έτσι έχει και το ανθρώπινο/συναισθηματικό στοιχείο. Διαχωρίζει την αρχιτεκτονική ζωγραφική σε τρεις τύπους, το αναφορικό σκίτσο, την προπαρασκευαστική μελέτη και τον τελικό πίνακα. Παρακάτω φαίνεται πως μπορούν να εφαρμοστούν οι σκέψεις του Graves στη διαδικασία εξαγωγής διορατικών λύσεων (insights) από τα δεδομένα:

### A. Δημιουργία σκίτσων

Μπορεί να ξεκινήσει ο διάλογος με τα δεδομένα φτιάχνοντας σκίτσα. Μπορούν να ζωγραφιστούν σκέψεις ή ιδέες για το επιθυμητό αποτέλεσμα.

### B. Προετοιμασία

Αυτό το στάδιο μπορεί να ξεκινήσει με μολύβι και χαρτί, αλλά θα χρειαστεί και λογισμικό για την ανάλυση και την οπτικοποίηση των δεδομένων. Μπορούν να χρησιμοποιηθούν τα παρακάτω για καταιγισμό ιδεών σχετικά με το πώς θα γίνει η προσέγγιση των δεδομένων:

- i. Φιλτράρισμα: Κόβεται ο θόρυβος και γίνεται επικέντρωση στο πιο ενδιαφέρον θέμα
- ii. Ταξινόμηση: Γίνεται ταξινόμηση των στοιχείων ανάλογα με την σημαντικότητά τους

- iii. Ομαδοποίηση και διαχωρισμός: Σύνοψη των στοιχείων και διαχωρισμός σε διαφορετικές ομάδες.
- iv. Οπτικοποίηση: Δίνεται ζωή στα δεδομένα

Οι αρχιτέκτονες πηγαίνουν μπρος πίσω από δισδιάστατα σε τρισδιάστατα σχέδια και παρά το ότι ένα σχέδιο μπορεί να είναι κατάλληλο και βάσιμο από μόνο του, ζωντανεύει μόνο όταν γίνεται πραγματικότητα. Στον κόσμο των δεδομένων συμβαίνει το ίδιο όταν αναμασώνται τα νούμερα. Μερικές φορές είναι ανακριβή, λάθος ή μη ολοκληρωμένα. Εδώ είναι το σημείο όπου η εργασία θα πρέπει να επιβεβαιώνεται συνεχώς με πραγματικά δεδομένα.

#### Γ.Οπτικοποίηση δεδομένων

Η οπτικοποίηση μπορεί να βοηθήσει στην μεταμόρφωση των στοιχείων σε χρήσιμες πληροφορίες. Όπως είπε και ο μαθηματικός John Tukey, «οι αριθμοί περιγράφουν τις αναμενόμενες τιμές, ενώ τα γραφήματα τις μη αναμενόμενες». Αυτό που θέλει να πει είναι ότι η πραγματική δύναμη της οπτικοποίησης των δεδομένων συμβάλλει στο να φανούν ευρήματα με έναν ξεκάθαρο και κατανοητό τρόπο.

Οι παρακάτω αρχές συμβάλλουν στην οπτικοποίηση που οδηγεί σε διορατικότητα (insights).

- i. Η αισθητική μετράει: Καθώς οι άνθρωποι προσελκύονται από τη φύση τους από την ομορφιά, γιατί να μην φτιαχτεί ένα γράφημα όσο πιο όμορφο γίνεται; Η αισθητική μπορεί επίσης να βοηθήσει στην πιο εύκολη κατανόηση της δράσης που θα πρέπει να γίνει τελικά.
- ii. Επικέντρωση στις τάσεις, όχι στα νούμερα: Οι καλύτερες διορατικές λύσεις (insights) προέρχονται από τις τάσεις, όχι από τα νούμερα, και ειδικά από αυτές που αλλάζουν κατεύθυνση.
- iii. Σύγκριση χρονικών διαστημάτων: Εξέταση διαφορετικών χρονικών διαστημάτων, όπως βδομάδα με βδομάδα, μήνα με μήνα, καλοκαίρι με καλοκαίρι. Εδώ βέβαια χρειάζεται και προσοχή στα διαστήματα που θα συγκριθούν, δεν μπορούμε για παράδειγμα να συγκρίνουμε τον Μάρτιο με τον Φεβρουάριο, γιατί ο δεύτερος έχει λιγότερες μέρες από τον πρώτο.
- iv. Διερεύνηση σημαντικών συσχετίσεων: Συχνά, οι πιο ενδιαφέρουσες ανακαλύψεις στην ανάλυση των δεδομένων είναι η συσχέτιση των μεταβλητών.



- v. Εξέταση των δεδομένων από διαφορετικές οπτικές γωνίες: Επειδή ένα άτομο δεν μπορεί να δει τα πάντα, καλό θα είναι να έχει και τη γνώμη της υπόλοιπης ομάδας.
- vi. Σκέψη: Πάντα θα πρέπει τα δεδομένα να αναλύονται από δύο οπτικές γωνίες. Για παράδειγμα, μπορεί με τα ίδια δεδομένα να γίνουν πολλαπλά διαγράμματα, ώστε να επιλεχθούν εκείνα που θα συμβάλλουν στην αφήγηση μίας ιστορίας με ακρίβεια.

### 3. Από τα δεδομένα στη δράση

Η μετατροπή των δεδομένων σε χρήσιμες διορατικές λύσεις (insights), μπορεί να είναι μία μη δομημένη διαδικασία. Ωστόσο, υπάρχουν διάφορες τεχνικές που βοηθούν τους επαγγελματίες σε αυτό. Οι μαρκετίερς θα πρέπει να δημιουργήσουν οδηγίες, παρόμοιες με το πρόγραμμα Blueprint των αρχιτεκτόνων, ώστε να δουλεύουν αποτελεσματικά με τα δεδομένα.

#### 5.3.6 Πηγαίνοντας πέρα από τα δεδομένα: Μετατροπή των δεδομένων σε διορατικές λύσεις (insights) και αξία [7]

Σύμφωνα με το άρθρο της KPMG “Going beyond the data: Turning data from insights into value”, που αφορά στα αποτελέσματα μίας έρευνας, η οποία πραγματοποιήθηκε ανάμεσα σε περισσότερα από 800 στελέχη με στόχο τα βασικά θέματα που καθορίζουν τις προτάσεις αύξησης της αξίας, οι οποίες προκύπτουν από τα δεδομένα και την ανάλυσή τους, τα βασικά συμπεράσματα είναι:

##### 1. Πρέπει να γίνεται καθορισμός των θεμάτων που αντιμετωπίζονται

Οι εταιρίες έχουν συνήθως ένα πρόβλημα, δεν ξέρουν ποιο αναλυτικό μοντέλο θα πρέπει να «τρέξουν» για να το λύσουν και πολλές φορές δεν ξέρουν καν ότι υπάρχει αυτή η λύση. Αυτό που πρέπει να κάνει το άτομο που καλείται να ερμηνεύσει τα δεδομένα είναι να βρει τα πιο σημαντικά σημεία μέσα στο πρόβλημα του πελάτη, στα οποία θα πρέπει να επικεντρωθεί το αναλυτικό μοντέλο.

##### 2. Η διορατικότητα (insights) δεν έρχεται μόνο από τα δεδομένα

Θα πρέπει να υπάρξει διατμηματική συνεργασία, ώστε να εντοπιστούν τα θέματα και οι προκλήσεις που αντιμετωπίζει μία εταιρεία.

### 3. Ο αναλυτής πρέπει να πάει πέρα από συγκεκριμένες λύσεις

Τα συγκεκριμένα έργα αναλυτικών μοντέλων δεν θα πρέπει να πραγματεύονται μία συγκεκριμένη λύση αλλά να είναι κομμάτι μίας ολοκληρωμένης επιχειρησιακής στρατηγικής και να στηρίζουν επενδυτικές αποφάσεις. Σύμφωνα με τον Christian Rast, Global Head of Data and Analytics της KPMG, η ανάπτυξη και η αύξηση της παραγωγικότητας είναι αυτά που προσθέτουν αξία σε έναν οργανισμό.

### 4. Καθορισμός της αξίας

Η αλήθεια είναι ότι η αξία διαφέρει ανάλογα με το θέμα το οποίο καλείται κάποιος να λύσει. Είναι διαφορετική αν χρειάζεται να μειωθεί το κόστος, διαφορετική όταν είναι να διαχειριστεί ο κίνδυνος και διαφορετική όταν χρειάζεται να βελτιωθεί η εμπειρία του πελάτη από ένα προϊόν ή μία υπηρεσία.

### 5. Επικέντρωση στους πελάτες

Θα πρέπει ο αναλυτής να σκεφτεί τρόπους για το πώς θα χρησιμοποιηθούν νέοι τύποι δεδομένων και αλγορίθμων ώστε να αυτοματοποιηθεί η λήψη αποφάσεων και να έχει ο πελάτης καλύτερη εξυπηρέτηση.

### 6. Πραγματοποίηση των σωστών ερωτήσεων

Δεν πρέπει να πραγματοποιείται ένα αναλυτικό μοντέλο μόνο και μόνο για να πραγματοποιηθεί. Θα πρέπει να μπουν σε προτεραιότητα οι διορατικές λύσεις (insights) και να κατανοηθεί η πιθανή αξία που έχουν αυτές, όχι μόνο σε όρους «του τι θα κερδίσει ο καθένας», αλλά και σε όρους ταχύτητας και πολυπλοκότητας.

### 7. Μέτρηση της επιτυχίας

Χρήση των επιτυχιών για την χρηματοδότηση περισσότερων έργων και μεταφορά της εμπειρίας και της γνώσης σε όλον τον οργανισμό.

### 8. Εμπλοκή των ενδιαφερόμενων μερών από τα πρώιμα στάδια ενός έργου

Πληροφόρηση του οργανισμού και των επενδυτών για την αξία των αναλυτικών μοντέλων, που τη βλέπουν όλο και πιο πολύ σαν μία στρατηγική μεταμόρφωσης και όχι απλά σαν έναν τρόπο για μεγαλύτερη διορατικότητα (insights) στα υπάρχοντα επιχειρηματικά προβλήματα.

## 6. Συμπεράσματα – Προτάσεις για περαιτέρω έρευνα

Τα Μεγάλα Δεδομένα και η ανάλυσή τους μπορούν να συμβάλλουν στη λήψη αποφάσεων η οποία θα βασίζεται σε δεδομένα, ώστε οι εταιρίες να αυξάνουν την αξία τους με την εκλογίκευση του κόστους, με την αύξηση της αποτελεσματικότητας, με την αύξηση της παραγωγικότητας ή με την διαχείριση του κινδύνου. Σύμφωνα με την IBM [27] τα δεδομένα αποτελούν τη νέα βάση για τη δημιουργία ανταγωνιστικού πλεονεκτήματος, την μετατροπή της διορατικότητας σε καινοτομία, στη δημιουργία νέων αποτελεσματικών επιχειρηματικών διαδικασιών, σε στρατηγικές αποφάσεις. Επιπλέον, δίνουν τη δυνατότητα στις εταιρίες να έχουν όλα τα δεδομένα και όχι μόνο ένα δείγμα για να ανακαλύψουν νέες τάσεις που πιθανό να μην ήταν ορατές πιο πριν και να εκμεταλλευτούν αυτές τις τάσεις για να καλύψουν με πιο αποτελεσματικό τρόπο τις ανάγκες των πελατών τους είτε δημιουργώντας ένα νέο προϊόν/υπηρεσία είτε βελτιώνοντας ένα υπάρχον προϊόν ή μία υπηρεσία.

Καθώς οι εταιρίες μπορούν να έχουν όλα τα δεδομένα και όχι μόνο ένα δείγμα του πληθυσμού, αυτό σχεδόν μηδενίζει τον κίνδυνο του στατιστικού σφάλματος και συμβάλει σημαντικά στην ακρίβεια των συμπερασμάτων. Παράλληλα, τα αναλυτικά μοντέλα υπάρχει η δυνατότητα να τροφοδοτούνται συνεχώς με νέα δεδομένα και επιτρέπουν την συνεχή παρακολούθηση των τάσεων ή τουλάχιστον σε τακτά χρονικά διαστήματα και πιο συχνά σε σχέση με τις παραδοσιακές μεθόδους έρευνας. Έτσι, οι αποφάσεις που λαμβάνονται με τη χρήση αναλυτικών μοντέλων είναι περισσότερο ακριβείς και λαμβάνονται με μεγαλύτερη ταχύτητα. Μία πρόταση για περαιτέρω έρευνα θα μπορούσε να είναι πως οι εταιρίες θα εξοικειωθούν με τη χρήση των αναλυτικών μοντέλων και θα τα χρησιμοποιούν στην καθημερινότητά τους για λήψη αποφάσεων.

Παράλληλα, η ανάλυση των Μεγάλων Δεδομένων δίνει τη δυνατότητα για ανάλυση κειμένου ή εικόνων ή αδόμητων δεδομένων γενικότερα τα οποία μέχρι πρόσφατα δεν ήταν εφικτό να αναλυθούν. Μεγάλο μέρος των δεδομένων που παράγονται γενικά είναι άχρηστο χωρίς την εφαρμογή κατάλληλων αναλυτικών μοντέλων. Σύμφωνα με την International Data Corporation (IDC), τα αναλυτικά μοντέλα εμπλουτισμένων μέσων (βίντεο, ήχος, εικόνες) υπολογίζεται ότι θα τριπλασιαστούν το 2015 και θα αναδειχθούν ως μία βασική κινητήρια δύναμη για επενδύσεις σε τεχνολογία για Μεγάλα Δεδομένα. Αυτά τα δεδομένα απαιτούν εξελιγμένα εργαλεία ανάλυσης. Άρα, λοιπόν μία πρόταση

για περαιτέρω έρευνα αποτελεί η μελέτη των εργαλείων που θα αναλύουν αδόμητα δεδομένα.

Δύο ακόμα τάσεις που χρήζουν περαιτέρω διερεύνησης αποτελούν τα Μεγάλα Δεδομένα στο σύννεφο και το Cognitive Computing [27]. Η International Data Corporation (IDC) προβλέπει αύξηση των δαπανών σε λύσεις ανάλυσης Μεγάλων Δεδομένων που βασίζονται σε σύννεφο κατά τη διάρκεια των επόμενων πέντε ετών. Από την άλλη πλευρά το Cognitive Computing αποτελεί μία τεχνολογία που χρησιμοποιεί επεξεργασία φυσικής γλώσσας και μηχανικής μάθησης για να βοηθήσει τους ανθρώπους και τις μηχανές να αλληλεπιδρούν με φυσικό τρόπο. Εξατομικευμένες εφαρμογές που θα χρησιμοποιούν cognitive computing θα βοηθάνε τους καταναλωτές να ψωνίζουν ρούχα, να επιλέγουν ένα μπουκάλι κρασί ή ακόμα και να δημιουργούν μια νέα συνταγή.

Μία άλλη τάση που χρήζει περαιτέρω διερεύνησης αποτελεί το Διαδίκτυο των Πραγμάτων, το οποίο αναμένεται να αυξηθεί σημαντικά τα επόμενα χρόνια και να οδηγήσει πολλούς οργανισμούς στην πρώτη χρήση των αναλυτικών μοντέλων συνεχούς ροής (streaming analytics). Τα αναλυτικά μοντέλα συνεχούς ροής και σε πραγματικό χρόνο απαιτούν επιστήμονες δεδομένων και ειδικούς θεμάτων που θα αναζητούν επαναλαμβανόμενα μοτίβα. Αναλυτικά εργαλεία και τεχνικές που θα έχουν τη δυνατότητα να διαχειριστούν τον τεράστιο όγκο των δομημένων και αδόμητων δεδομένων που παράγονται από το Διαδίκτυο των Πραγμάτων έρχονται σταδιακά στο φως αφού τα παραδοσιακά αναλυτικά μοντέλα δεν μπορούν να δουλέψουν καλά με τα θορυβώδη, αναλογικά, υψηλής ταχύτητας δεδομένα που παράγονται από τους αισθητήρες.

## Βιβλιογραφία

1. Laney, Douglas, 2001, "3D Data Management: Controlling Data Volume, Velocity and Variety", Gartner.
2. Beyer, Mark, 2011, "Gartner Says Solving 'Big Data' Challenge Involves More Than Just Managing Volumes of Data", Gartner.
3. Villanova University, 2015, "What is Big Data", University Alliance.
4. Henry K, 2014, "Big Data Paris 2014", P.Delort, Paris, France.
5. Billings S.A, 2013, "Nonlinear System Identification: NARMAX Methods in the Time Frequency and Spatio-Temporal Domains", Wiley.
6. Henry K, 2013, "Big Data Paris 2013", P.Delort, Paris, France.
7. Delort P, 2013, "Big Data car Low-Density Data? La faible densité en information comme facteur discriminant".
8. De Mauro, Andrea; Greco, Marco; Grimaldi, Michele, 2015, "what is big data? A consensual definition and a review of key research topics", AIP Conference Proceedings 1644: 97–104, doi:10.1063/1.4907823.
9. Bailey, Ken, 1994, "Numerical Taxonomy and Cluster Analysis (Typologies and Taxonomies)", p. 34. ISBN 9780803952591.
10. Jewels, Berman, 2013, "Principles of Big Data 1<sup>st</sup> edition", Morgan Kaufmann.
11. APICS (2012), 2012, "APICS 2012 Big Data Insights and Innovations Executive Summary".
12. Philip Carter, 2011, "Big Data Analytics: Future Architectures, Skills and Roadmaps for the CIO", SAS.
13. Gobble, M. M, 2013, "Big Data: The Next Big Thing in Innovation", Research Technology Management, 56(1), 64-66.
14. Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C. and Byers, A. H, 2011, "Big data: The next frontier for innovation, competition, and productivity", McKinsey Global Institute.
15. Mayer-Schönberger, V. and Cukier, K, 2013, "big data: A revolution that will transform how we live, work, and think", Houghton Mifflin Harcourt.
16. Russom, P, 2011, "Big data analytics", TDWI Best Practices Report (Fourth Quarter).
17. Sagiroglu, S. and Sinanc, D, 2013, "Big data: A review", IEEE International Conference on Collaboration Technologies and Systems, San Diego, USA.
18. Strawn, G. O, 2012, "Scientific Research: How Many Paradigms?", EDUCAUSE Review, 47(3), 26-34.

19. Δημήτρης Μαρκόπουλος, 2014, Hans Jessen: Ο Δανός που μετέφερε στην Ελλάδα την επίλεκτη ομάδα του ομίλου του, newmoney.gr
20. IRI, Introduction to Client Value Program (CVP) – Corporate Training & Development
21. IRI, Business Insights – Corporate Training & Development
22. KPMG International, 2015, Going beyond the data: turning data from insights into value, KPMG International Cooperative
23. Chris Meyer, Tim McGuire, Maher Masri, Abdul Wahab Shaikh, 2013, Four Steps To Turn Big Data Into Action, Forbes
24. Laura Patterson, 2013, 5 Ways to Turn Big Data Into Insight and Action, Online Marketing Institute
25. Daniel Waisberg, 2014, From Data to Insights: The Blueprint for Your Business, Google
26. Thomas H.Davenport and D.J.Patil, October 2012, issue, Data Scientist: The Sexiest Job of the 21st Century, Harvard Business Review
27. Hema Reddy, September 2015, Big data trends: The top eight analytics lessons for business, IBM Big Data & Analytics Hub
28. Eileen McNulty, May 2014, Understanding Big Data: The seven v's, Dataconomy
29. Diya Soubra, July 2012, The 3vs that define big data, Data Science Central
30. Data Science in Wikipedia, Retrieved November 2015, from [https://en.wikipedia.org/wiki/Data\\_science](https://en.wikipedia.org/wiki/Data_science)
31. Machine Learning in Wikipedia, Retrieved November 2015, from [https://en.wikipedia.org/wiki/Machine\\_learning](https://en.wikipedia.org/wiki/Machine_learning)
32. Barton Poulson, Retrieved November 2015, Techniques and Concepts of Big Data, Lynda.com
33. Harlan Harris, Sean Murphy, Marck Vaisman, June 2013, Analyzing the Analyzers - An Introspective Survey of Data Scientists and Their Work, O'Reilly

## Άλλες Πηγές

<http://www.journals.elsevier.com/journal-of-business-research/call-for-papers/special-issue-on-big-data-and-analytics-in-technology/>

<http://searchbusinessanalytics.techtarget.com/definition/big-data-analytics>

<http://searchbusinessanalytics.techtarget.com/definition/big-data-analytics>

<http://searchenginewatch.com/sew/how-to/2289574/big-data-big-trouble-how-to-avoid-5-data-analysis-pitfalls>

<http://www.sas.com/resources/asset/five-big-data-challenges-article.pdf>

<http://www.gsma.com/membership/wp-content/uploads/2013/07/The-Top-Challenges-of-Big-Data-Analytics.pdf>

<http://www.computerworld.com/article/2690856/big-data/8-big-trends-in-big-data-analytics.html>

<https://blogs.saphana.com/2015/06/16/new-sap-hana-sps10/>

<http://www.cio.com/article/2881201/data-analytics/8-analytics-trends-to-watch-in-2015.html>

<http://www.ibmbigdatahub.com/blog/big-data-trends-top-eight-analytics-lessons-business>

[http://www.sas.com/en\\_us/insights/big-data/what-is-big-data.html?gclid=CjwKEAiA7MWyBRDpi5TFqqmm6hMSJAD6GLeAe\\_9y5aarjk9RTzt8FTDjHCNq0\\_aXluZSq49UO\\_vQThoClvTw\\_wcB](http://www.sas.com/en_us/insights/big-data/what-is-big-data.html?gclid=CjwKEAiA7MWyBRDpi5TFqqmm6hMSJAD6GLeAe_9y5aarjk9RTzt8FTDjHCNq0_aXluZSq49UO_vQThoClvTw_wcB)