

Πτυχιακή Εργασία

Ανακάλυψη Γνώσης από Δεδομένα Κοινωνικών Δικτύων

Ανδρέου Γεώργιος

ΑΕΜ:2334

Επιβλέπων Καθηγητής:

κ. Βλαχάβας Ιωάννης



Αριστοτέλειο
Πανεπιστήμιο
Θεσσαλονίκης
Σχολή Θετικών Επιστημών
Τμήμα Πληροφορικής

Θεσσαλονίκη 2019

Περίληψη

Η ραγδαία ανάπτυξη των κοινωνικών δικτύων έχει οδηγήσει σε ριζικές αλλαγές στην καθημερινότητα των ανθρώπων αλλά και των επιχειρήσεων. Πλέον αποτελούν αναπόσπαστο κομμάτι σε κάθε σύγχρονη επιχείρηση και είναι το βασικότερο μέσο προώθησης και επικοινωνίας με το αγοραστικό κοινό. Για αυτό οι πληροφορίες που υπάρχουν στα κοινωνικά δίκτυα σχετικά με τους ακόλουθους και το αγοραστικό κοινό μίας εταιρίας αποτελούν καθοριστικό κομμάτι για την στρατηγική που θα ακολουθήσει η εταιρία και για τις αποφάσεις που θα πάρει. Το Facebook είναι το κυρίαρχο κοινωνικό δίκτυο στις μέρες μας καθώς σχεδόν κάθε επιχείρηση διατηρεί μία σελίδα εκεί. Στα πλαίσια της εργασίας καλούμαστε να εξάγουμε γνώση μέσα από τα δεδομένα που μας προσφέρει το Facebook για μία σελίδα και τους ακόλουθους της. Εξάγουμε τα δεδομένα αυτά και τα χρησιμοποιούμε για αποκτήσουμε μια καλύτερη εικόνα για την κατάσταση και την απήχηση της σελίδας, για να ανακαλύψουμε συσχετίσεις μεταξύ των δεδομένων και να τα συγκρίνουμε με τα δεδομένα της ελληνικής στατιστικής αρχής και τέλος να καταλήξουμε σε συμπεράσματα σχετικά με το βέλτιστο target group της επιχείρησης για μία διαφημιστική καμπάνια.

To repository της εφαρμογής: <https://github.com/gioandreou/thesis-auth>

Abstract

The rapid growth of social networks has lead to radical changes not only in humans' daily life but also in business operations. Social networks constitute an integral part of every contemporary business and the leading medium of promotion and consumer communication. In this context, the information that exists about the consumers and the followers of a corporation's social network plays an important role in the effectiveness of business strategies and in decision-making processes. It is widely accepted that Facebook is the leading social network nowadays as the vast majority of business has page there. In the context of this thesis, we are called to extract knowledge through the data that Facebook give us for a page and its followers. The extraction of this data provides us a better insight into the impact and effectiveness of the page, the discovery of correlations between the data and their comparison to the data of the Hellenic statistical authority and finally helps drawing conclusions about the optimal target group at advertising campaigns.

Repository of the application: <https://github.com/gioandreou/thesis-auth>

Ευχαριστίες

Σε αυτό το σημείο θα ήθελα να εκφράσω τις θερμές μου ευχαριστίες σε όλους εκείνους που συνέβαλλαν και βοήθησαν στην εκπόνηση αυτής της πτυχιακής εργασίας.

Θα ήθελα πρωτίστως να ευχαριστήσω θερμά τον καθηγητή κύριο Ιωάννη Βλαχάβα για την εμπιστοσύνη που μου έδειξε και μου ανέθεσε την εργασία αυτή αλλά και για την συνεχόμενη στήριξη και καθοδήγηση με τις γνώσεις του σε όλους αυτούς τους μήνες μέχρι να ολοκληρωθεί η εργασία. Επίσης τον ευχαριστώ θερμά για τον χρόνο που αφιέρωσε για τις συναντήσεις μας, στις ερευνητικές συζητήσεις μας αλλά και τις προτροπές του να με φέρει σε επικοινωνία με άλλους καθηγητές για επιπλέον υποστήριξη.

Ένα μεγάλο ευχαριστώ στους συμφοιτητές και φίλους που με στήριξαν και μου πρόσφεραν τις προγραμματιστικές γνώσεις τους όποτε τις χρειαζόμουν. Παράλληλα θα ήθελα να ευχαριστήσω φίλους και άτομα εκτός του ακαδημαϊκού περιβάλλοντος και της πληροφορικής για την στήριξη που μου πρόσφεραν και που ήταν πάντα πρόθυμοι να με βοηθήσουν και να μου εκφράσουν την αντικειμενική τους γνώμη για την εφαρμογή που δημιουργήθηκε.

Τέλος θα ήθελα να δώσω το μεγαλύτερο ευχαριστώ στην οικογένεια μου, στους γονείς μου και στον αδερφό μου συγκεκριμένα που ήταν πάντα δίπλα μου, με συνεχόμενη πίστη στις δυνατότητες μου και στα όνειρα μου.

Θεσσαλονίκη, Μάρτιος 2019

Ανδρέου Γεώργιος

Περιεχόμενα

Περίληψη	3
Abstract	4
Ευχαριστίες	5
Κεφάλαιο 1	11
Κεφάλαιο 2 : Κοινωνικά Μέσα Δικτύωσης	13
2.1 Εισαγωγή	13
2.2 Ορισμός Κοινωνικών Δικτύων	14
2.3 Χαρακτηριστικά των Κοινωνικών Δικτύων	15
2.4 Χρησιμότητα των Κοινωνικών Δικτύων	17
2.5 Τα Δημοφιλέστερα Social Networks	18
2.5.1. Facebook	19
2.5.2 Twitter	20
2.5.3 LinkedIn	21
2.5.4. Google+	22
2.5.5. Instagram	22
2.6 Η Ιστορία των Κοινωνικών Δικτύων	24
2.8 Στατιστικά Χρήσης των Κοινωνικών Δικτύων	27
2.9 Social Media Marketing(SMM)	30
2.10 Στρατηγικές Social Media Marketing	31
Σύνοψη Κεφαλαίου	32
Κεφάλαιο 3 : Τα Δεδομένα και η Επεξεργασία τους	33
3.1 Εισαγωγή	33
3.1.1 Η Επιστήμη των Δεδομένων	34
3.1.2 Big Data	35
3.1.3 GDPR Κανονισμός Προστασίας Δεδομένων	37
3.2 Εξόρυξη Δεδομένων (Data Mining)	39
3.2.1 Εξόρυξη Δεδομένων στα Κοινωνικά Δίκτυα	41
3.2.2. Facebook Graph API	43
3.3 Ανάλυση των Δεδομένων	44
3.4 Μηχανική Μάθηση	46

3.4.1 Τι ορίζουμε μηχανική μάθηση	47
3.4.2. Επιβλεπόμενη μάθηση (Supervised Learning)	48
3.4.3 Μη επιβλεπόμενη Μάθηση (Unsupervised Learning)	52
3.4.4 Ενισχυτική Μάθηση (Reinforcement Learning)	53
Σύνοψη Κεφαλαίου	53
Κεφάλαιο 4 : Ανάλυση Δεδομένων Κοινωνικών Δικτύων	55
4.1 Δεδομένα από Facebook Page	55
4.1.1 Facebook For Developers	55
4.1.2. Δεδομένα και Μετρικές που Χρησιμοποιήθηκαν	57
4.1.3. Δημογραφικά Δεδομένα	61
4.1.4. Αποθήκευση Δημογραφικών Δεδομένων	62
4.2 Υλοποίηση Εφαρμογής και Ανάπτυξη Κώδικα	63
4.2.1. Βιβλιοθήκες και Packages	64
4.2.2. Αρχεία Python	64
4.3 Δεδομένα ΕΛΣΤΑΤ	70
4.3.1 Δεδομένα που Χρησιμοποιήθηκαν	70
4.3.2. Συσχέτιση Δεδομένων ΕΛΣΤΑΤ	72
4.4. Γραφήματα Ploty – Dash	73
4.4.1. Το Dashboard της Σελίδας	73
Σύνοψη Κεφαλαίου	83
Κεφάλαιο 5 : Συμπεράσματα	85
Βιβλιογραφία	89

Λίστα Εικόνων

Εικόνα 2.1. Γράφημα με τους λόγους χρήσης των social networks.	19
Εικόνα 2.2. Landing Page του Friendster και διαφημιστική αφίσα του CompuServe	25
Εικόνα 2.3. Η πρωταρχική όψη της κεντρικής σελίδας του Facebook	26
Εικόνα 2.4. % των Αμερικανών ενηλίκων και χρηστών Ίντερνετ	28
Εικόνα 2.5. Γραφήματα κοινωνικών δικτύων σε σχέση με Φύλο-Ηλικία-Εισόδημα-Εκπαίδευση.	29
Εικόνα 3.1. Η συσχέτιση της επιστήμης των δεδομένων με άλλα πεδία.	34
Εικόνα 3.2. Τα Βασικά βήματα της εξόρυξης δεδομένων	41
Εικόνα 3.3. Η εφαρμογή του Graph API με ένα αίτημα για τα ενδιαφέροντα των φίλων.	44
Εικόνα 3.5. Η πρώτη πρόβλεψη του συστήματος για $h(x) = 12 + 0.2x$	49
Εικόνα 3.6. Η βέλτιστη συνάρτηση που παράγει το σύστημα.	50
Εικόνα 3.7. Γραφική αναπαράσταση της συνάρτησης κόστους.	51
Εικόνα 3.8. Γραφική αναπαράσταση της συνάρτησης $g(z)$.	52
Εικόνα 4.1. To Dashboard στην αρχική σελίδα στο Facebook for Developers	55
Εικόνα 4.2. To Facebook Graph API Explorer.	56
Εικόνα 4.3. Οι άδειες για τα δεδομένα που επιθυμούμε να πάρουμε	56
Εικόνα 4.4. To JSON Dictionary που επιστρέφει το request.	61
Εικόνα 4.5 Αριστερά το αρχείο που αποθηκεύονται ηλικίες και δεξιά οι πόλεις των ακόλουθων.	62
Εικόνα 4.6. Στατιστικά στοιχεία για την σελίδα και τις αναρτήσεις της σελίδας	66
Εικόνα 4.7. Στατιστικά στοιχεία για τις ηλικίες των ακόλουθων της σελίδας.	66
Εικόνα 4.8. Τα αποτελέσματα που εξάγονται από την συσχέτιση τα δεδομένα της ΕΛΣΤΑΤ.	67
Εικόνα 4.9. Γραφήματα με πληροφορίες της σελίδας μέσω του matplotlib	67
Εικόνα 4.10. Η εφαρμογή Flask με τα γραφήματα της ΕΛΣΤΑΤ	68
Εικόνα 4.11 Τα callbacks στις μεθόδους του plotly_elstat.py	68
Εικόνα 4.12 Τα δημογραφικά δεδομένα της ΕΛΣΤΑΤ στην εφαρμογή.	69
Εικόνα 4.13 To dashboard της σελίδας με τα 3 πρώτα γραφήματα.	70
Εικόνα 4.14 Το αρχείο της ΕΛΣΤΑΤ για την οικογενειακή κατάσταση	71

Εικόνα 4.15 Τα τροποποιημένα αρχεία για την οικογενειακή κατάσταση και τη κατάσταση ασχολίας των Ελλήνων.	72
Εικόνα 4.16 Το γράφημα για τις ηλικίες των ακολούθων	74
Εικόνα 4.17 Το bar chart με τις περιοχές των ακολούθων.	75
Εικόνα 4.18 Η δυνατότητα να προσκαλέσει ο διαχειριστής νέα άτομα στη σελίδα.	76
Εικόνα 4.19 Το line chart με τις πληροφορίες για την σελίδα	76
Εικόνα 4.20 Το line chart με τις πληροφορίες για τις αναρτήσεις της σελίδας	77
Εικόνα 4.21 Το διαδραστικό line chart πόλεων	78
Εικόνα 4.22 Το διαδραστικό line chart χωρών	78
Εικόνα 4.23 Ο παγκόσμιος χάρτης ακολούθων	79
Εικόνα 4.24 Το γράφημα με τις πληροφορίες για κάθε ανάρτηση	80
Εικόνα 4.25 Οι στατιστικές κάθε ανάρτησης	81
Εικόνα 4.26 Το διάστημα όπου η συγκεκριμένη ανάρτηση είναι ζωντανή (alive)	81
Εικόνα 4.27 Το γράφημα ηλικιών-φύλου στις συγκεκριμένες ημερομηνίες κάθε ανάρτησης.	82
Εικόνα 4.28 Το line chart για τα γεωγραφικά διαμερίσματα σε συγκεκριμένες ημερομηνίες	82
Εικόνα 4.29 Το target group με τη μεγαλύτερη απήχηση για κάθε ανάρτηση	83
Εικόνα 5.1 Ορίζουμε το budget της διαφημιζόμενης ανάρτησης.	85
Εικόνα 5.2 Τα αποτελέσματα της διαφημιζόμενης ανάρτησης.	86
Εικόνα 5.3 Η νέα διαφημιζόμενη ανάρτηση στο συγκεκριμένο target group που ορίσαμε.	87
Εικόνα 5.3 Τα αποτελέσματα της νέας διαφημιζόμενης ανάρτησης	87

Κεφάλαιο 1

Αντικείμενο της πτυχιακής εργασίας

Στη σημερινή κοινωνία τα κοινωνικά δίκτυα και οι εφαρμογές τους παίζουν ολοένα και σημαντικότερο ρόλο στην καθημερινότητα μας, στον τρόπο που επικοινωνούμε, μέχρι και στον τρόπο που λαμβάνουμε αποφάσεις. Μακριά από αυτήν την ζώνη επιφροής των κοινωνικών δικτύων δεν θα μπορούσαν να μείνουν οι εταιρίες και επιχειρήσεις.

Τα κοινωνικά δίκτυα αποτελούν πλέον ένα πολύ σημαντικό εργαλείο για τις επιχειρήσεις καθώς δίνουν την δυνατότητα τις άμεσης επικοινωνίας με τους πελάτες και το αγοραστικό κοινό τους. Επιπλέον τα κοινωνικά δίκτυα προσφέρουν γρήγορες και άμεσες πληροφορίες όσο αφορά τις τάσεις και τις προτιμήσεις του κοινού, κάτι το οποίο ήταν αρκετά χρονοβόρο και δαπανηρό στο παρελθόν μέσα από ερωτηματολόγια και έρευνες αγοράς.

Πλέον σχεδόν κάθε επιχείρηση είναι ενεργή στα κοινωνικά δίκτυα με αρκετούς ακόλουθους στις σελίδες τους. Όταν μία σελίδα σε ένα κοινωνικό δίκτυο π.χ. Facebook έχει αρκετούς ακόλουθους, τότε μπορούμε να χρησιμοποιήσουμε τα δεδομένα που έχουμε για αυτούς ώστε να βελτιώσουμε τόσο την εικόνα της επιχείρησης στα κοινωνικά δίκτυα όσο και να βελτιστοποιήσουμε τις διαφημιστικές μας καμπάνιες ή την προώθηση νέων προϊόντων.

Συχνά όμως υπάρχει ένα πρόβλημα με την πληθώρα πληροφοριών που μπορεί να έχουμε σχετικά με την σελίδα και τους ακόλουθους μας. Τα δεδομένα τις περισσότερες φορές είναι τόσα πολλά ή περίπλοκα δομημένα ώστε να μην μπορούμε να τα ερμηνεύσουμε σωστά και να πάρουμε σωστές αποφάσεις.

Σκοπός της πτυχιακής εργασίας

Το πρόβλημα αυτό καλείται να λύσει η πτυχιακή εργασία μέσα από την εφαρμογή που αναπτύχθηκε. Μέσα από μία δοκιμαστική σελίδα που δημιουργήσαμε στο Facebook, παίρνουμε δεδομένα για τη σελίδα και τους ακόλουθους της και τα χρησιμοποιούμε ώστε να εξάγουμε χρήσιμες πληροφορίες για αυτήν.

Έχοντας δημογραφικά δεδομένα για τους ακόλουθους της σελίδας αλλά και πληροφορίες για την απήχηση της σελίδας και των αναρτήσεων της, μπορούμε μέσω της ανάλυσης αυτών των δεδομένων να δούμε σημαντικές λεπτομέρειες και συσχετίσεις μεταξύ των δεδομένων, οι οποίες δεν θα ήταν αλλιώς ορατές.

Χρησιμοποιώντας επίσης ορισμένα δεδομένα της Ελληνικής στατιστικής έχουμε μια καλύτερη εικόνα και στοιχεία για τους ακόλουθους της σελίδας μας. Όσο περισσότερες πληροφορίες

έχει μία επιχείρησή για τους ακόλουθους της στα κοινωνικά δίκτυα και το εν δυνάμη αγοραστικό κοινό της, τόσο καλύτερα μπορεί να εκτιμήσει το αγοραστικό της κοινό και να τους προσφέρει προϊόντα που έχουν ανάγκη.

Στη δικιά μας περίπτωση περιοριζόμαστε στον εντοπισμό μίας ομάδας ατόμων ή ενός αγοραστικού κοινού στο οποίο μία εταιρία κατά τη διάρκεια της διαφημιστικής της καμπάνιας θα έχει τα καλύτερα αποτελέσματα με το μικρότερο κόστος.

Δομή της πτυχιακής εργασίας

Το **2^ο Κεφάλαιο** παρουσιάζει αρχικά μια χρονολογική αναδρομή στην ιστορία των κοινωνικών δικτύων και στις μορφές που έχουν και τα χαρακτηριστικά που τα διακατέχουν μέχρι σήμερα. Παρουσιάζονται επίσης πληροφορίες για τα σημαντικότερα κοινωνικά δίκτυα που υπάρχουν σήμερα καθώς και την επίδραση των κοινωνικών δικτύων στον χώρο του μάρκετινγκ μέσα από το social media marketing.

Το **3^ο Κεφάλαιο** περιγράφει τα δεδομένα και τις εφαρμογές τους. Ειδικότερα παρουσιάζονται πληροφορίες για την επιστήμη δεδομένων, τα μεγάλα δεδομένα (Big Data) και γίνεται αναφορά στον νέο κανονισμό προστασίας δεδομένων (GDPR). Στη συνέχεια αναφέρεται το κομμάτι της εξόρυξης δεδομένων στα κοινωνικά δίκτυα και πιο συγκεκριμένα στο Facebook, μέσα από το Graph API. Τέλος ακολουθεί ο ορισμός της μηχανικής μάθησης και τα είδη της μηχανικής μάθησης με ορισμένα παραδείγματα.

Το **4^ο Κεφάλαιο** περιγράφει λεπτομερώς την ανάπτυξη της εφαρμογής για την εξαγωγή γνώσης από μία σελίδα Facebook. Περιγράφονται αναλυτικά τα στάδια προγραμματισμού της εφαρμογής, τα δεδομένα που παίρνουμε από το Facebook και την Ελληνική στατιστική αρχή, τις προγραμματιστικές βιβλιοθήκες που χρησιμοποιούμε, τα αρχεία που παράγονται καθώς και τη διαδραστική παρουσίαση των δεδομένων.

Το **5^ο Κεφάλαιο** περιέχει τα αποτελέσματα τα οποία παράγουμε με την χρήση της εφαρμογής σε 2 διαφημιζόμενες αναρτήσεις της σελίδας μας και τα συμπεράσματα στα οποία καταλήγουμε για τη χρησιμότητα της εφαρμογής μας.

To repository της εφαρμογής: <https://github.com/gioandreou/thesis-auth>

Κεφάλαιο 2

Κοινωνικά Μέσα Δικτύωσης

2.1 Εισαγωγή

Ο 21ος αιώνας μπορεί εύκολα να χαρακτηρισθεί ως τεχνολογικός αιώνας και ο αιώνας της επικοινωνίας, μιας και τα νέα δεδομένα υποδηλώνουν μια ραγδαία αύξηση της χρήσης του ηλεκτρονικού υπολογιστή, του διαδικτύου και κατ' επέκταση των κοινωνικών δικτύων σε όλα τα επίπεδα της καθημερινότητας. Η έμφυτη ανάγκη επικοινωνίας των ανθρώπων έχει καταστήσει τα κοινωνικά δίκτυα ιδιαίτερα δημοφιλή στο ευρύ κοινό, με αποτέλεσμα την ανάπτυξη αυτών μέρα με μέρα, ακολουθώντας τον σύγχρονο τρόπο ζωής.

Η κοινωνική δικτύωση σήμερα είναι ιδιαίτερα δημοφιλής στην online μορφή της, δηλαδή σε απευθείας σύνδεση στο Διαδίκτυο. Το Διαδίκτυο πλέον είναι γεμάτο με εκατομμύρια άτομα που επιθυμούν να γνωρίσουν άλλους ανθρώπους, να αναπτύξουν φιλίες ή επαγγελματικές συμμαχίες, να αναζητήσουν εργασία, να σχηματίσουν ομάδες και να ανταλλάξουν γρήγορα και εύκολα πληροφορίες και εμπειρίες σχετικά με τα χόμπι και τα ενδιαφέροντά τους. Όσον αφορά την επιχειρηματική στρατηγική, τα social media χρησιμοποιούνται ως μέσα δημιουργίας εταιρικής εικόνας, ενημέρωσης, επικοινωνίας και ανάπτυξης σχέσεων με τους πελάτες. Τα κοινωνικά δίκτυα έχουν ιδιαιτέρως μεγάλη απήχηση σε ένα ευρύ κοινό, επομένως η άποψη ότι ζούμε την επανάσταση των Social Media δεν είναι καθόλου τυχαία ή υπερβολική.

Σύμφωνα με την πανελλήνια έρευνα του εργαστηρίου Ηλεκτρονικού Εμπορίου, οι Έλληνες χαρακτηρίζονται ως τακτικοί και έμπειροι χρήστες, αφού το 50% δήλωσε πως χρησιμοποιεί τα social media πάνω από 3 χρόνια. Το 68% των Ελλήνων επισκέπτεται σελίδες κοινωνικής δικτύωσης και το 60% επισκέπτεται blogs και φόρουμ. Παράλληλα το 50% αναζητά πληροφορίες σε ιστοσελίδες με αξιολογήσεις χρηστών. Από την πλευρά των επιχειρήσεων, τουλάχιστον η μία στις τρεις χρησιμοποιεί κάποια πλατφόρμα κοινωνικής δικτύωσης. Οι εταιρίες χρησιμοποιούν τα social media για να ενισχύσουν την παραγωγικότητα, την καινοτομία, τη φήμη και τη συνεργασία των εργαζομένων με την εταιρία. [1]

Όταν πρόκειται για online κοινωνική δικτύωση, χρησιμοποιούνται οι ιστοσελίδες γνωστές ως ιστότοποι κοινωνικής δικτύωσης. Οι ιστότοποι κοινωνικής δικτύωσης λειτουργούν ως online κοινότητες των χρηστών του Διαδικτύου. Ανάλογα με τα χαρακτηριστικά κάθε ιστότοπου, τα μέλη της online κοινότητας έχουν κοινά ενδιαφέροντα όπως: χόμπι, θρησκεία ή πολιτική. Η εγγραφή και πρόσβαση σε έναν ιστότοπο κοινωνικής δικτύωσης παρέχει στα μέλη του, άμεση και συνεχή

κοινωνικοποίηση. Αυτή η κοινωνικοποίηση μπορεί να περιλαμβάνει την ανάγνωση του προφίλ των άλλων μελών και συχνά την επικοινωνία μαζί τους.

Οι ιστοσελίδες κοινωνικής δικτύωσης ή ηλεκτρονικά κοινωνικά δίκτυα, έκαναν την εμφάνισή τους το 2002 με το Friendster. Αυτές οι ιστοσελίδες επιτρέπουν στα άτομα να παρουσιάσουν τους εαυτούς τους, να αναπτύξουν την κοινωνική τους δικτύωση, καθώς και να δημιουργήσουν ή να διατηρήσουν συνδέσεις με άλλους χρήστες. Η ιστοσελίδα κοινωνικής δικτύωσης όμως που ξεχώρισε ανάμεσα στις άλλες είναι το Facebook. Με 2.19 δισεκατομμύρια ενεργούς μηνιαίους χρήστες και 13% αύξηση στο πρώτο τρίμηνο του 2018 το Facebook είναι αναμφισβήτητα η δημοφιλέστερη ιστοσελίδα κοινωνικής δικτύωσης [2]

Επιπλέον, τα κοινωνικά δίκτυα χωρίζονται σε άμεσα και έμμεσα δίκτυα. Τα άμεσα δίκτυα παρέχουν πρόσβαση στο προφίλ ενός χρήστη χωρίς να χρειάζεται η συγκατάθεσή του. Ενώ στα έμμεσα δίκτυα είναι απαραίτητο ο χρήστης να επιτρέπει στον άλλον τη σύνδεση με το προφίλ του. Τα πρώτα χρησιμοποιούνται κυρίως για διαφήμιση και μάρκετινγκ και τα δεύτερα παράγουν τη word-of-mouth (από στόμα-σε-στόμα) επικοινωνία, που οδηγεί στη διαμόρφωση μιας συμπεριφοράς των καταναλωτών. [3]

2.2 Ορισμός Κοινωνικών Δικτύων

Σύμφωνα με την Wikipedia, ο όρος μέσα κοινωνικής δικτύωσης αναφέρεται στα μέσα αλληλεπίδρασης και επικοινωνίας ομάδων ανθρώπων μέσω διαδικτυακών κοινοτήτων. Τα online κοινωνικά δίκτυα ορίζονται ως web-based υπηρεσίες που επιτρέπουν τα άτομα να δημιουργήσουν ένα δημόσιο ή ημι-δημόσιο προφίλ μέσα σε ένα οριοθετημένο σύστημα, να επικοινωνήσουν με ένα πλήθος από άλλους χρήστες με τους οποίους μοιράζονται την ίδια μορφή σύνδεσης, τη λίστα των συνδέσεών τους, όσο και αυτών που φτιάχτηκαν από άλλους χρήστες μέσα στο σύστημα. Εν γένει, κοινωνική δικτύωση είναι η συγκέντρωση ή συμμετοχή των ατόμων σε συγκεκριμένες ομάδες. Τα κοινωνικά δίκτυα ορίζονται ως ένα σύνολο από actors (άνθρωποι, οργανισμοί ή άλλες κοινωνικές ομάδες) και ένα σύνολο από τις σχέσεις τους (φιλίες, δεσμοί, χρηματικές συναλλαγές κ.α.) ή την έλλειψη αυτών ανάμεσα στους actors. [1]

Οι όροι “social media” και “social network” συχνά ταυτίζονται κάτω από τον όρο “κοινωνική δικτύωση”. Ωστόσο, κατά τη μελέτη και των δύο, διαπιστώνεται σημαντική διαφοροποίηση μεταξύ τους. Σε μια πρώτη ανάγνωση, ο όρος «social media» αναφέρεται στα μέσα (εργαλεία) διαμοιρασμού της πληροφορίας, των δεδομένων και της επικοινωνίας στο κοινό, ενώ ο όρος «social networking» αναφέρεται στη δημιουργία και την αξιοποίηση κοινοτήτων για τη διασύνδεση ανθρώπων με κοινά ενδιαφέροντα. Δε θα ήταν λάθος να ειπωθεί ότι ο όρος «social media» αναφέρεται στα εργαλεία –

μέσα ενημέρωσης και κοινωνικής δικτύωσης, ενώ ο όρος «social networking» στη διαδικασία της κοινωνικής δικτύωσης. [1]

Ένας ακόμη πιο ακριβής ορισμός στα κοινωνικά μέσα δικτύωσης αναφέρει, πως ένα κοινωνικό δίκτυο είναι περιβάλλον κυβερνοχώρου που επιτρέπει σε έναν χρήστη να δομήσει το προφίλ του, να μοιραστεί κείμενο, εικόνα και συνδέσμους προς τα άλλα μέλη του ιστοτόπου μέσω εφαρμογών ή ομάδων σε αυτό. Έτσι το κοινωνικό δίκτυο δίνει τη δυνατότητα στον χρήστη να αυτοπαρουσιάζεται, να χτίζει και να διατηρεί σχέσεις με άλλους. [4]

2.3 Χαρακτηριστικά των Κοινωνικών Δικτύων

Τα μέσα κοινωνικής δικτύωσης, αποτελούν μέρος των εφαρμογών του Web 2.0. Ο κάθε χρήστης εισέρχεται ως μέλος και μπορεί να δημιουργήσει έναν εικονικό προσωπικό χώρο με τα πραγματικά του στοιχεία (ονοματεπώνυμο, τόπος κατοικίας, φωτογραφίες κλπ.), όπου επικοινωνεί με άλλους, μοιράζεται τις ιδέες του, δείχνει βίντεο και φωτογραφίες και κυρίως συνδέεται με άλλους. Η συσπείρωση των χρηστών γίνεται πλέον με βάση τους άλλους ανθρώπους, με τους οποίους συνδέονται και όχι με βάση τα ενδιαφέροντά τους. Κατά τη διαδικασία της συναναστροφής τους, οι χρήστες, αποδέχονται και μοιράζονται αξίες που διαμορφώνονται από τα χαρακτηριστικά των τεχνολογιών των μέσων κοινωνικής δικτύωσης. Τη συνεργασία και τη σύμπραξη, όπου κανείς δεν έχει τον έλεγχο πάνω στον άλλο, τη διάδοση μέσω της πρότασης και όχι της επιβολής, τη χρήση κατόπιν επιλογής, την ελεύθερη πρόσκληση για σύνδεση και την κοινή συνεισφορά.

Η διαδικασία κοινωνικοποίησης ενός κοινωνικού δικτύου Web 2.0 συγκριτικά με ένα παραδοσιακό Internet group Web 1.0, διαφέρει στο ότι, πλέον στην όλη διαδικασία συμπεριλαμβάνονται ενέργειες που καθορίζουν οι χρήστες με βάση την κουλτούρα και τις αξίες του group στο οποίο ταυτοποιούνται.

Τα παρακάτω είναι τα αντιπροσωπευτικά χαρακτηριστικά ενός Web 2.0 κοινωνικού δικτύου.

[5]

- Φιλίες:** Τα κοινωνικά δίκτυα δίνουν την δυνατότητα στους χρήστες έκφρασης άποψης και γνώμης και αυτό πιθανολογικά μπορεί να οδηγήσει σε φιλίες. Εναλλακτικά, ο χρήστης μπορεί να προσκαλέσει τους φίλους του να συμμετάσχουν σε αυτό το online group.
- Δημοκρατική Διαμέριση:** Τα κοινωνικά δίκτυα του Web 2.0 παρακινούν την ατομική προσφορά του καθενός ώστε να διαμοιραστεί σε ολόκληρο το δίκτυο.

- **Αξιοποίηση Συλλογικής Νοημοσύνης:** Το πλεονέκτημα των Web 2.0 ιστοσελίδων βασίζεται στην επικριτική μάζα από συμμετέχοντες. Επομένως το κλειδί στις δυναμικές των groups είναι οι κοινωνικές επιδράσεις από τις συνεισφορές των χρηστών.
- **Viral Promotion:** Τα κοινωνικά δίκτυα του Web 2.0 στηρίζονται στην προώθηση μέσω word-of-mouth. Αυτό γίνεται όταν ο χρήστης μιλά-διαφημίζει ένα site, μια υπηρεσία, μία εκδήλωση ή ένα προϊόν, μοιράζοντας στους άλλους τις θετικές του εμπειρίες.
- **Καινοτομία στην Ενσωμάτωση:** Μέσω της ενσωμάτωσης υπηρεσιών από διαφορετικά άτομα, οι κοινότητες του Web 2.0 μπορούν να δημιουργήσουν μια αξία που μπορεί να έχει οικονομικό όφελος.
- **Pull but not Push:** Τα συστήματα “Pull” αφήνουν τους ανθρώπους να φέρουν τις σχέσεις και τα περιεχόμενα εκεί που θέλουν, αντί να τους τα καθορίζει κάποιος άλλος και να ασκεί πίεση. Στα Web 2.0 social media οι χρήστες έχουν τον έλεγχο των συνομιλιών τους.
- **Συνεργασία, Σύμπραξη, αλλά όχι Έλεγχο:** Οι Web 2.0 εφαρμογές είναι δομημένες πάνω σε υπηρεσίες δεδομένων που συνεργάζονται. Επομένως δεν υπάρχει έλεγχος της χρήσης δεδομένων.

Σε μια άλλη έρευνα, που διεξάχθηκε από το Max Planck Institute και τα πανεπιστήμια Maryland και Rice, αναλύθηκαν τα χαρακτηριστικά τους μέσω κοινωνικών δικτύων και κατέληξαν σε 5 δομικά χαρακτηριστικά που υπάρχουν στα μέσα κοινωνικής δικτύωσης. [6]

- **User-based:** Αρχικά, πριν από τα social networks οι ιστοσελίδες λειτουργούσαν με βάση το περιεχόμενο που είχε ανεβάσει ένας χρήστης και διαβαζόταν από τους επισκέπτες. Η ροή των πληροφοριών ήταν προς μια κατεύθυνση και η κατεύθυνση των ενημερώσεων καθοριζόταν από τον webmaster ή τον συγγραφέα. Αντιθέτως τα κοινωνικά δίκτυα είναι δομημένα και χτισμένα από τους ίδιους τους χρήστες για τους ίδιους. Χωρίς αυτούς το δίκτυο θα ήταν ένας κενός χώρος. Αυτό είναι επίσης και το βασικό συστατικό που τα κάνει τόσο συναρπαστικά και ελκυστικά.
- **Αλληλεπίδραση:** Ένα άλλο βασικό χαρακτηριστικό των σύγχρονων κοινωνικών δικτύων είναι η διαδραστικότητα. Αυτό σημαίνει ότι ένα κοινωνικό δίκτυο δεν είναι απλώς μια συλλογή των δωματίων συνομιλίας και φόρουμς, αντιθέτως μάλιστα τα περισσότερα έχουν εφαρμογές παιχνιδιών κ.α. Μπορούν να προσαρμόζονται στις ανάγκες των χρηστών τους, παρέχοντας τους λειτουργίες και υπηρεσίες που χρησιμοποιούν και τους διευκολύνουν, όταν είναι συνδεδεμένοι στο διαδίκτυο.

- **Κοινότητα:** Όπως οι κοινότητες ή οι κοινωνικές ομάδες σε όλο τον κόσμο στηρίζονται στο γεγονός ότι τα μέλη κατέχουν κοινές πεποιθήσεις ή χόμπι, τα κοινωνικά δίκτυα λειτουργούν με βάση την ίδια αρχή. Οι χρήστες αποτελούν μέρος της κοινότητας και μπορούν να ανταλλάξουν απόψεις, ιδέες, φωτογραφίες κ.α.
- **Σχέσεις:** Σε αντίθεση με τους ιστότοπους του παρελθόντος, πλέον στα κοινωνικά δίκτυα αναπτύσσονται σχέσεις. Όσο μεγαλύτερος ο αριθμός των σχέσεων, φιλιών και ακολούθων που έχει ένας χρήστης τόσο πιο δημοφιλής είναι μέσα στο δίκτυο.
- **Συναίσθημα:** Βασικό χαρακτηριστικό των κοινωνικών δικτύων είναι ο συναίσθηματικός παράγοντας. Ενώ οι ιστοσελίδες επικεντρώνονται κατά κύριο λόγο στην παροχή πληροφοριών, το κοινωνικό δίκτυο παρέχει μια συναίσθηματική ασφάλεια στους χρήστες του, ότι οι φίλοι τους βρίσκονται σε κοντινή απόσταση και εύκολα μπορούν να επικοινωνήσουν μαζί τους.

2.4 Χρησιμότητα των Κοινωνικών Δικτύων

Σε πολλά από τα μεγάλα κοινωνικά δίκτυα, οι συμμετέχοντες δεν κάνουν απαραίτητα δικτύωση ή δεν επιζητούν να γνωρίσουν νέους ανθρώπους. Πρωτίστως επικοινωνούν με άτομα που ήδη βρίσκονται στον κοινωνικό τους δίκτυο. Μεταξύ των πιο κοινών χρήσεων των social media είναι οι εξής:

- **Σύνδεση με υπάρχοντα δίκτυα και ανάπτυξη φιλιών:** Οι νέοι συνήθως χρησιμοποιούν τα κοινωνικά δίκτυα για να επικοινωνήσουν και να κοινωνικοποιηθούν με τις επαφές τους. Πολύ συχνά κάποια άτομα χρησιμοποιούν τα social media για να συλλέξουν επαφές με σκοπό να επιδείξουν τη δημοτικότητά τους.
- **Online Αντιπροσώπευση:** Οι υπηρεσίες κοινωνικής δικτύωσης παρέχουν χώρους για τα μέλη ώστε να μπορούν να δημιουργούν το online profile τους και να αντιπροσωπεύονται από αυτό στο ευρύτερο δίκτυο.
- **Ενημέρωση-Εύρεση πληροφοριών:** Εκτός από το να παραμένουν ενήμεροι για τις δραστηριότητες των άλλων χρηστών, οι χρήστες μπορούν επίσης να ενημερωθούν για την επικαιρότητα και για τα δρώμενα της πόλης τους, αν είναι συνδεδεμένοι με τις κατάλληλες ομάδες χρηστών και ενημερωτικών δικτύων.
- **Δημιουργία και Ανέβασμα Περιεχομένου:** Περιεχόμενο με τη μορφή φωτογραφίας, βίντεο, κειμένου ή ήχου μπορεί να ανέβει στο προφίλ του χρήστη. Παράλληλα μπορεί να γίνει και

διανομή εξωτερικού περιεχομένου, όπως για παράδειγμα ένας σύνδεσμος για περιεχόμενο που φιλοξενείται κάπου αλλού.

- **Διαγωνισμοί και Ψηφοφορίες:** Πλέον είναι αρκετά εύκολο για έναν χρήστη να θέσει ερωτήματα προς τον διαδικτυακό κοινωνικό του κύκλο και να λάβει γρήγορα αρκετές απαντήσεις. Η μορφή των απαντήσεων δεν περιορίζεται σε κείμενο μόνο αλλά μπορεί να έχει μία από τις προαναφερθέντες μορφές περιεχομένου.
- **Μηνύματα-Δημόσια & Ιδιωτικά:** Οι περισσότερες από τις υπηρεσίες κοινωνικής δικτύωσης υποστηρίζουν την αποστολή και λήψη δημοσίων και ιδιωτικών μηνυμάτων ανάμεσα σε χρήστες. Σε ορισμένα κοινωνικά δίκτυα υπάρχει και η δυνατότητα προσωρινού ιδιωτικού μηνύματος που διαρκεί μερικά μόνο λεπτά, από την στιγμή που θα διαβαστεί.

Σύμφωνα με μια δημοσκόπηση που διεξάχθηκε από την Global Web Index σε άτομα ηλικίας 16-64 με θέμα “Ποιοι οι κυριότεροι λόγοι χρήσης των social media ?” διαπιστώθηκε, πως το 55% των ερωτηθέντων χρησιμοποιεί τα κοινωνικά δίκτυα για να μένει ενήμερος με τις δραστηριότητες των φίλων τους. Αυτό το συμπέρασμα εδραιώνει το λεγόμενο “παθητική δικτύωση”, στην οποία ο χρήστης χρησιμοποιεί τα social media κυρίως για να καταναλώσει περιεχόμενο, παρά να δημιουργήσει ο ίδιος. Οι κυριότεροι λόγοι για την χρήση των social media, σύμφωνα με την δημοσκόπηση, είναι εξής:

1. Να μένεις ενήμερος με την δραστηριότητα των φίλων.
2. Να ενημερώνεσαι για την επικαιρότητα και τα νέα.
3. Να γεμίσεις τον ελεύθερο χρόνο.
4. Να βρίσκεις αστείο και ψυχαγωγικό περιεχόμενο.
5. Να μοιράζεσαι γνώμες.
6. Να μοιράζεσαι φωτογραφίες ή βίντεο με άλλους.
7. Επειδή ήδη οι φίλοι χρησιμοποιούν τα κοινωνικά δίκτυα.
8. Δικτύωση με άλλα άτομα, πέρα από το κύκλο.
9. Να γνωρίζεις καινούργια άτομα.
10. Να μοιράζεσαι πληροφορίες για την καθημερινότητα. [7]

2.5 Τα Δημοφιλέστερα Social Networks

Ο χώρος των social network είναι ένα ρευστό και συνεχές εξελισσόμενο οικοσύστημα, το οποίο επηρεάζεται από την καθημερινότητα και τις εκάστοτε τάσεις. Τα social networks έχουν αλλάξει κατά πολύ από τις πρώτες μέρες της εμφάνισής τους μέχρι σήμερα και θα συνεχίσουν να

αλλάζουν με την πάροδο του χρόνου. Έχουμε περάσει από την εποχή του Myspace σε μία εποχή που κυριαρχείται κατά κράτος από το Facebook και από κάποια άλλα κοινωνικά δίκτυα.

2.5.1. Facebook

To Facebook ξεκίνησε τη λειτουργία του, το 2004, όταν ιδρύθηκε από τον φοιτητή του Harvard Mark Zuckerberg. Αρχικός του στόχος, ήταν να αποτελέσει έναν χώρο κοινωνικής δικτύωσης, στον οποίο θα είχαν δικαίωμα συμμετοχής μόνο οι φοιτητές του Harvard. Οι χρήστες της υπηρεσίας μπορούσαν να εγγραφούν δωρεάν και να δημιουργήσουν προσωπικό λογαριασμό χρήστη. Κατόπιν, μπορούσαν να επικοινωνούν μέσω μηνυμάτων με τις επαφές τους, να προσθέτουν και να ανανεώνουν τις προσωπικές πληροφορίες τους και να δημιουργούν ομάδες για οποιοδήποτε θέμα. Όλοι είχαν πρόσβαση να συμμετάσχουν σε δίκτυα που σχετίζονται με το πανεπιστήμιο, τις θέσεις απασχόλησης ή τις γεωγραφικές περιοχές.



Εικόνα 2.1. Γράφημα με τους λόγους χρήσης των social networks.

Το όνομα της υπηρεσίας προέρχονταν από τα έγγραφα παρουσίασης των μελών πανεπιστημιακών κοινοτήτων ορισμένων Αμερικάνικων κολεγίων και προπαρασκευαστικών σχολείων που χρησιμοποιούσαν οι νεοεισερχόμενοι σπουδαστές για να γνωριστούν μεταξύ τους. Το 2005 το δικαίωμα πρόσβασης επεκτάθηκε σε μαθητές συγκεκριμένων λυκείων και μέλη ορισμένων

μαθητικών κοινοτήτων, ενώ το 2006 η υπηρεσία έγινε προσβάσιμη σε κάθε άνθρωπο του πλανήτη που η ηλικία του ξεπερνούσε τα 13 χρόνια.

Το Facebook το 2013 είχε πάνω από 1 δισεκατομμύριο ενεργούς χρήστες. Επίσης, το Facebook είναι ένα από τα δημοφιλέστερα sites για ανέβασμα φωτογραφιών με πάνω από 14 εκατομμύρια φωτογραφίες καθημερινά. Το Facebook είναι ένας καλός τρόπος δικτύωσης με φίλους και γνωστούς. Παρά το ότι ενέχει κινδύνους (κυρίως για παραβίαση προσωπικών δεδομένων), ο προσεκτικός χρήστης δεν έχει πρόβλημα. Το Facebook ακόμα παρέχει παιχνίδια και υπάρχει η δυνατότητα ανεβάσματος φωτογραφιών και βίντεο.

Το Facebook έχει υποστεί κριτική και έχει κατηγορηθεί για θέματα ασφαλείας, προστασίας της ιδιωτικότητας και για τις πολιτικές απόψεις των ιδρυτών τους. Ωστόσο, η συγκεκριμένη υπηρεσία παραμένει το πιο διάσημο social network σε πολλές αγγλόφωνες -και όχι μόνο- χώρες.

Είναι το κορυφαίο σε επισκεψιμότητα μέσο κοινωνικής δικτύωσης, καθώς πέρασε το Myspace τον Απρίλιο του 2008 και είναι πρώτο σε επισκεψιμότητα, αντίστοιχα, έχει την υψηλότερη επισκεψιμότητα ανάμεσα στα μέσα κοινωνικής δικτύωσης, έχοντας περίπου 2 δισεκατομμύρια μηνιαίους ενεργούς χρήστες και πάνω από 1 δισεκατομμύριο χρήστες που συνδέονται καθημερινά.
[8] [9]

Χρησιμοποιεί αρκετές τεχνολογίες, εργαλεία και γλώσσες προγραμματισμού στη δομή της. Στο κομμάτι του Front-End, οι servers λειτουργούν σε LAMP (Linux, Apache, MySQL, και PHP). Στο Back-End χρησιμοποιούνται οι C++, JAVA, Python και Erlang με την εξής φιλοσοφία:

1. Δημιούργησε μια υπηρεσία αν χρειαστεί.
2. Δημιούργησε ένα Framework/toolset για ευκολότερη δημιουργία υπηρεσιών.
3. Χρησιμοποίησε την κατάλληλη προγραμματιστική γλώσσα για το κάθε έργο.

Για την αποθήκευση δεδομένων χρησιμοποιείται η Cassandra (database), η οποία χειρίζεται μεγάλους όγκους δεδομένων ανάμεσα σε διάφορους servers.[10]

2.5.2 Twitter

Το Twitter είναι ένας ιστοχώρος κοινωνικής δικτύωσης που επιτρέπει στους χρήστες του να στέλνουν και να διαβάζουν σύντομα μηνύματα (μέχρι 280 χαρακτήρες), τα οποία ονομάζονται tweets. Τα μηνύματα μπορούν να αναγνωστούν και από μη συνδεδεμένους χρήστες, αλλά μόνο οι συνδεδεμένοι μπορούν να δημοσιεύσουν κείμενα.

Δημιουργήθηκε τον Μάρτιο 2006 από τους Jack Dorsey, Noah Glass, Biz Stone και Evan Williams, το οποίο ξεκίνησε τον Ιούλιο της ίδιας χρονιάς. Η υπηρεσία γρήγορα έλαβε αρκετή

δημοσιότητα. Το 2012, περισσότεροι από 100 εκατομμύρια χρήστες δημοσίευσαν 340 εκατομμύρια tweets σε μία μέρα, και η υπηρεσία διαχειρίστηκε κατά μέσο 1,6 δισεκατομμύρια αναζητήσεις σε μία μέρα. Το 2013, ήταν μία από τις 10 ιστοσελίδες με τη μεγαλύτερη επισκεψιμότητα και περιγράφηκε ως “το SMS του Ίντερνετ”. Την ημέρα των αμερικανικών εκλογών του 2016, το Twitter ήταν το μεγαλύτερο μέσο άμεσης ενημέρωσης.

Τον Μάρτιο του 2018, το Twitter βαθμολογήθηκε ως το 12o σε επισκεψιμότητα ιστότοπο σύμφωνα με την ανάλυση της Alexa. Το Twitter είχε μια ετήσια τάξη αύξησης 1.382%, αυξάνοντας από 475.000 επισκέπτες το Φεβρουάριο του 2008 σε 7 εκατομμύρια τον Φεβρουάριο του 2009. Η παγκόσμια βάση δεδομένων του αρίθμησε 328 εκατομμύρια χρήστες το 2017.

Το Twitter βασίζεται αρκετά σε open-source λογισμικό με το Twitter Web interface να χρησιμοποιεί το Ruby on Rails framework. Αρχικά τα tweets αποθηκεύονταν σε βάσεις MySQL, αλλά λόγω του μεγάλου όγκου ήταν επιτακτική ανάγκη να αλλάξει το σύστημα. Η αλλαγή έγινε σταδιακά και χρησιμοποιήθηκε η Scala για να γραφτεί το νέο software. Η αλλαγή αυτή προσέφερε μία ώθηση της τάξης 10 φορές πάνω στην απόδοση.

2.5.3 LinkedIn

Το LinkedIn αποτελεί ένα δικτυακό τόπο κοινωνικής δικτύωσης που εστιάζει κυρίως σε επιχειρηματικές δραστηριότητες. Δίνει τη δυνατότητα στους χρήστες να δημιουργήσουν ένα δίκτυο από υπάρχουσες και νέες επαγγελματικές επαφές. Όμως, προχωρά και σε άλλες δράσεις. Συγκεκριμένα, όταν μια επιχείρηση προσθέτει μια νέα σύνδεση, οι συνδέσεις αυτής και οι συνδέσεις αυτών των συνδέσεων επίσης προστίθενται στο δίκτυο. Αυτό προσφέρει σ' έναν επαγγελματία μεγάλο πεδίο επαγγελματικών επαφών.

Σήμερα θεωρείται ο πιο επιτυχημένος ιστοχώρος επαγγελματικής κοινωνικής δικτύωσης παγκοσμίως, μετρώντας περισσότερους από 300 εκατομμύρια εγγεγραμμένους χρήστες σε περισσότερες από 200 χώρες. Σύμφωνα με την Πανελλήνια ετήσια Έρευνα του 2012 για τη συμπεριφορά των χρηστών των Μέσων Κοινωνικής Δικτύωσης από το Εργαστήριο Ηλεκτρονικού Εμπορίου- ELTRUN, του Οικονομικού Πανεπιστημίου Αθηνών, πάνω από 511.291 χρήστες του LinkedIn βρίσκονται στην Ελλάδα. [11]

Για τις επιχειρήσεις που αναζητούν τρόπο να προωθήσουν τις υπηρεσίες τους, το LinkedIn είναι ένα αρκετά χρήσιμο εργαλείο. Μέσα από αυτό δημιουργείται ένα δίκτυο συνδέσεων, όπου οι επιχειρηματίες μπορούν να βελτιώσουν το προφίλ τους μέσα σε αυτή την κοινότητα. Χρησιμοποιώντας ένα ισχυρό portfolio και κατάλογο επιτευγμάτων εξασφαλίζει ο χρήστης ότι αυτοί που ενδιαφέρονται για τα συγκεκριμένα “χαρακτηριστικά” και “επιτεύγματα” θα τον προτιμήσουν.

Ένα ακόμα χαρακτηριστικό του LinkedIn είναι οι λεγόμενες recommendations. Όταν προσφέρει μια επιχείρηση μια υπηρεσία σε κάποιο άλλο άτομο, μέλος της κοινότητας, μπορεί να δώσει ένα σχόλιο για την προσφορά των υπηρεσιών. Τα σχόλια αυτά είναι ένα αρκετό σημαντικό κομμάτι και εργαλείο για την επιλογή ενός παρόχου υπηρεσιών, καθώς μπορούν να διαμορφώσουν την εταιρική γνώμη κάθε εταιρίας.

To LinkedIn έχει ένα δικό του μοντέλο API που χρησιμοποιεί, το Rest.li. Κυρίως χρησιμοποιείται JAVA, αλλά έχει και αρκετές εφαρμογές που χρησιμοποιούν Python, Ruby, Node.js, and C++.

2.5.4. Google+

To Google plus (Google+) είναι ένα κοινωνικό δίκτυο από την μεγαλύτερη εταιρεία διαδικτυακών υπηρεσιών το οποίο φέρει και το όνομα της. Η Google μετά από κάποιες αποτυχημένες προσπάθειες κοινωνικών δικτύων (π.χ. Buzz, Wave κτλ.) δημιούργησε το Google plus με στόχο να εδραιωθεί και να ανταγωνιστεί άλλα μεγάλα κοινωνικά δίκτυα όπως το Facebook και το Twitter.

To Google+ ξεκίνησε τον Ιούνιο του 2011. Ο χρήστης έχει την δυνατότητα όπως και στα άλλα κοινωνικά δίκτυα να ανεβάσει περιεχόμενο και ενδιαφέροντα σε κοινότητες και σε γκρουπ ατόμων, τα λεγόμενα “Circles”. Από την στιγμή που έχει δημιουργηθεί ένα circle, ο χρήστης μπορεί να μοιράσει περιεχόμενο αποκλειστικά μόνο στα άτομα μέσα στον “κύκλο”. Στην αρχική σελίδα του κάθε χρήστη, η οποία ονομάζεται “Stream”, εμφανίζονται στον καθένα τα τελευταία posts από τους “κύκλους”, που συμμετέχει ο χρήστης.

Για την άμεση επικοινωνία μεταξύ χρηστών χρησιμοποιείται το Hangouts, ένα άλλο προϊόν της Google με μεγάλη επιτυχία. Στο Hangouts, οι χρήστες μπορούν να στείλουν εύκολα και γρήγορα μηνύματα αλλά και να δημιουργήσουν ομαδικές συνομιλίες.

To Google+ είναι μία τυπική Google Web App. Χρησιμοποιεί Java servlets για τον κώδικα του Server και JavaScript για την browser-side του User Interface. Κυρίως όμως βασίζεται στο Framework της Google, το Closure, συμπεριλαμβανομένου και του JavaScript complier και του template system. To Back-End έχει χτιστεί κυρίως πάνω σε Bigtable, Colossus/GFS και MapReduce. [12]

2.5.5. Instagram

To Instagram είναι μια δωρεάν εφαρμογή κοινωνικής δικτύωσης που επιτρέπει την επεξεργασία και την κοινοποίηση φωτογραφιών και βίντεο. Οι χρήστες μπορούν να μοιράζονται φωτογραφίες και βίντεο με τους ακολούθους τους ή μια συγκεκριμένη ομάδα ατόμων. Ο χρήστης μπορεί να σχολιάσει και να δηλώσει ότι του αρέσει μια δημοσίευση.

Η εφαρμογή δημιουργήθηκε από δύο απόφοιτους του Πανεπιστημίου του Stanford, τους Κέβιν Σίστρομ και Μάικ Κρίγκερ και ξεκίνησε τον Οκτώβριο του 2010. Μόλις δύο μήνες αργότερα, το Δεκέμβριο του 2010, ο αριθμός των εγγεγραμμένων χρηστών έφτασε το 1.000.000. Σήμερα η εφαρμογή μετράει 20 δισεκατομμύρια φωτογραφίες από όλο τον κόσμο και 800 εκατομμύρια ενεργούς χρήστες. Το όνομά της προέρχεται από τον συνδυασμό της λέξης **Instant** (στιγμιαίο) και **telegram** (τηλεγράφημα). Το 2012 η εφαρμογή αγοράστηκε από το Facebook, προς ένα 1 δισεκατομμύρια δολάρια Η.Π.Α.

Οι φωτογραφίες έχουν τετράγωνο σχήμα, παρόμοιο με το καρέ των φωτογραφικών μηχανών Kodak Instamatic και Polaroid. Οι χρήστες μπορούν να τις επεξεργαστούν και να διαμορφώσουν το τελικό αποτέλεσμα μέσω ειδικά σχεδιασμένων φίλτρων και ειδικών εργαλείων επεξεργασίας. Η δυνατότητα εγγραφής και διαμοιρασμού βίντεο ενσωματώθηκε στο Instagram τον Ιούνιο του 2013. Η εγγραφή βίντεο υποστηρίζεται προς το παρόν μόνο στις εφαρμογές iPhone και Android. Η μέγιστη διάρκεια του βίντεο είναι το ένα λεπτό με ελάχιστο χρόνο τα τρία δευτερόλεπτα.

Τελευταίες προσθήκες στα χαρακτηριστικά και στα εργαλεία της εφαρμογής είναι τα IGTV, Instagram Direct και Instagram Stories.

- Το IGTV είναι μία εφαρμογή προβολής κάθετων βίντεο, τα οποία δεν μπορούν να ξεπερνούν τα 10 λεπτά σε διάρκεια και τα 650MB.
- Το Instagram Direct είναι μια λειτουργία που επιτρέπει στους χρήστες να επικοινωνούν με ιδιωτικά μηνύματα. Από τον Σεπτέμβριο του 2015, μέσα στα Direct μηνύματα υπάρχει το conversation threading, το οποίο επιτρέπει την κοινοποίηση τοποθεσίας, hastags, και προφίλ άλλων χρηστών.
- Τον Αύγουστο του 2016, το Instagram εισήγαγε τα Stories, μια λειτουργία που επιτρέπει τη λήψη φωτογραφιών, προσθήκη φίλτρων και “αυτοκόλλητων” επαυξημένης πραγματικότητας. Όλα αυτά ανεβαίνουν στον ειδικό χώρο των Stories. Οι φωτογραφίες που ανεβαίνουν διαγράφονται μετά από ένα 24ωρο.

Το Instagram για να εξυπηρετήσει τον μεγάλο όγκο δεδομένων σε mobile συσκευές χρησιμοποιεί React Native για εύκολο μοίρασμα κώδικα. Επίσης χρησιμοποιούνται Amazon web services για μεγάλες υπολογιστικές επιδόσεις. Πιο συγκεκριμένα Amazon EC2, Route 53, S3, EBS, Gunicorn. Το Ubuntu Linux 11.04 είναι το λειτουργικό σύστημα και οι servers χρησιμοποιούν το Django framework σε Python. Τα δεδομένα αποθηκεύονται μέσω της PostgreSQL στις βάσεις δεδομένων.

2.6 Η Ιστορία των Κοινωνικών Δικτύων

Όλα ξεκίνησαν με τα BBS (Bulletin Board Systems), τα οποία ήταν online χώροι συνάθροισης ατόμων, που επέτρεπαν στους χρήστες να επικοινωνούν με ένα κεντρικό σύστημα από το οποίο κατέβαζαν αρχεία ή και παιχνίδια. Τα BBS διαχειριζόντουσαν κυρίως από χομπίστες, οι οποίοι φρόντιζαν αρκετά για τις κοινωνικές πτυχές και τη φύση του περιεχομένου του πρότζεκτ τους. Τα BBS είχαν τρομερή απήχηση την δεκαετία του 80 και μέχρι και τα μέσα της δεκαετίας του 90 ήταν ακόμα αρκετά δημοφιλή.

Ένα από αυτά και αρκετά γνωστό ήταν το CompuServe, μια υπηρεσία που ξεκίνησε το 1970 ως μέσο επικοινωνίας υπολογιστών καθαρά αφοσιωμένο για επαγγελματίες. Στη συνέχεια εξαπλώθηκε και σε δημόσια χρήση το 1988. Το CompuServe επέτρεπε στα μέλη του να μοιράζονται αρχεία και πρόσβαση σε ειδήσεις και εκδηλώσεις. Μπορούσε κανείς να συμμετάσχει στα χιλιάδες φόρουμς που υπήρχαν για να συζητήσει κάθε τι που ήταν επίκαιρο και θέμα της ημέρας. Αυτά τα φόρουμς έχτισαν τον δρόμο για τα μοντέρνα δίκτυα όπως τα ξέρουμε σήμερα.

Αν όμως θέλουμε να ονομάσουμε τον πραγματικό πρόδρομο των σημερινών κοινωνικών δικτύων, τότε αυτό ήταν το AOL (American Online). Για πολλούς το AOL ήταν το “ίντερνετ πριν το ίντερνετ”, και σίγουρα οι κοινότητες που είχαν δημιουργήσει τα μέλη του ήταν το πιο συναρπαστικό χαρακτηριστικό του.

Το 1995 όμως ήρθε το site το οποίο ταιριάζει στο σύγχρονο ορισμό του κοινωνικού δικτύου. Το Classmates.com ήταν το πρώτο site που απαντούσε στην “Με ποιον μπορώ να συνδεθώ, με τον οποίο κάποτε είμασταν συμμαθητές”. Η ιδέα της ψηφιακής επανένωσης έγινε γρήγορα μεγάλη επιτυχία και μέχρι σήμερα αριθμεί 57 εκ. μέλη

Το 2002 ο τομέας της κοινωνικής δικτύωσης έκανε το μεγάλο βήμα με το Friendster. Το Friendster χρησιμοποιούσε έναν βαθμό διαχώρισης-ομαδοποίησης με τον οποίο φτιαχνόντουσαν “κύκλοι φίλων”. Σύμφωνα με την ιδέα, ότι μία πλούσια διαδικυακή κοινότητα μπορεί να υπάρξει μόνο μεταξύ ατόμων που έχουν κοινά ενδιαφέροντα. Το Friendster διασφάλιζε ότι υπήρχε τρόπος να ανακαλυφθούν αυτοί οι δεσμοί κοινών ενδιαφερόντων. Μέσα σε έναν χρόνο από την αφετηρία του το Friendster αριθμούσε 3 εκ. εγγεγραμμένους χρήστες και αμέτρητο ενδιαφέρον για επένδυση.

Έναν χρόνο αργότερα, το 2003, το LinkedIn ήρθε στην επιφάνεια με μία αρκετά πιο σοβαρή και επαγγελματική προσέγγιση στον χώρο των κοινωνικών δικτύων. Αντί να αποτελεί έναν "παιδότοπο" για παλιούς συμμαθητές και εφήβους, το LinkedIn ήταν μία πηγή για κοινωνικοποίηση αφοσιωμένη σε επαγγελματίες. Οι επαφές μέσα στο LinkedIn αναφέρονται ως "συνδέσεις - connections". Σήμερα περισσότεροι από 297εκ. χρήστες χρησιμοποιούν το LinkedIn καθημερινά.

The image shows two side-by-side screenshots from the early 2000s. On the left is the 'Friendster beta' landing page, featuring a logo with a smiley face icon and the word 'friendster'. Below it, the text 'Friendster Beta' and 'The new way to meet people'. A central paragraph describes Friendster as an online community connecting people through networks of friends for dating or making new friends. It lists three ways to use the service: meeting new people to date, making new friends, and helping friends meet new people. To the right is a CompuServe advertisement titled 'Welcome to someday.' It features a man sitting at a desk with a computer monitor displaying a graph, and a woman standing beside him. The text discusses the future of home computing, mentioning shopping, banking, news analysis, stock performance, email, and playing Bridge. It also promotes CompuServe's Information Service, noting its availability in major US cities and its connection to various computers and databases. The CompuServe logo and address are at the bottom.

Εικόνα 2.2. Landing Page του Friendster και διαφημιστική αφίσα του CompuServe

Το MySpace ξεκίνησε επίσης το 2003. Αν και σήμερα δεν κατέχει τα ηνία στα κοινωνικά μέσα, στις αγγλόφωνες χώρες το MySpace ήταν κάποτε το μεγάλο χιτ, ο απόλυτος κυρίαρχος. Αυτό συνέβει λόγω της προσέγγισης εφήβων, ένα ηλικιακό γκρουπ κλειδί για την επιτυχία, και της συσχέτισης αυτού του γκρουπ με την μουσική, τα βίντεο και ένα funky-alternative περιβάλλον.

Ανεπτυγμένο από φοιτητές του Harvard, όπως αρκετά networking sites, το Facebook προοριζόταν να γίνει ο απόλυτος κυρίαρχος στα κοινωνικά δίκτυα. Το Facebook ξεκίνησε το 2004 ως αποκλειστικό μέσο των φοιτητών του Harvard, μέχρι και το 2006 όπου έγινε πλέον ανοιχτό σε όλους. Ακόμα και τότε το Facebook, θεωρούνταν πως είχε απίστευτες δυνατότητες, σε τέτοιο βαθμό, όπου ο συνιδρυτής του PayPal, Peter Thiel, επένδυσε 10 εκ. δολάρια μόνο και μόνο για να δει το Facebook να πετυχαίνει. Το μυστικό της επιτυχίας του Facebook είναι θέμα μεγάλου debate. Πολλοί μιλάνε για την ευκολία στην χρήση του, άλλοι για το "πιασάρικο-catchy" όνομά του. Όπως και να έχει, είναι κοινώς αποδεκτό πως το Facebook, προάγει και προωθεί την ειλικρίνεια. Οι άνθρωποι απολαμβάνουν να είναι ο εαυτός τους και να παρουσιάζουν την πραγματική πλευρά τους εκεί έξω.

Επίσης σημαντικός παράγοντας στην επιτυχία του Facebook ήταν το Facebook Platform το 2007, ένα ανοιχτό API, που επέτρεπε σε third-party developers να δημιουργήσουν εφαρμογές για το Facebook. Μετά από λίγο καιρό το API είχε τέτοια επιτυχία, όπου το Facebook έφτιαξε το Facebook App Store για να οργανώσει όλες αυτές τις εφαρμογές. Ένα άλλο χαρακτηριστικό της επιτυχίας ήταν το κουμπί “Like”. To like σαν εικονίδιο, σαν όρος, σαν έννοια ξέφυγε από τα όρια του Facebook και πλέον υπάρχει παντού στο ίντερνετ.



Εικόνα 2.3. Η πρωταρχική όψη της κεντρικής σελίδας του Facebook

Αντιλαμβάνοντας τη δυναμική των κοινωνικών δικτύων η Google αποφάσισε να ξεκινήσει και αυτή την δικιά της πλατφόρμα κοινωνικής δικτύωσης το 2007, το Google+. Διέφερε από το Facebook και το Twitter, σε έναν βαθμό όπου δεν ήταν ακριβώς ένα κοινωνικό δίκτυο, αλλά ένα κοινωνικό “στρώμα-πλευρά” στην όλη Google experience. Η Google δημιούργησε αρκετές καλές εντυπώσεις με την υπηρεσία “Hangouts”, η οποία επιτρέπει στους χρήστες να επικοινωνούν με video chats με τους φίλους τους. Μέσα σε 4 εβδομάδες, το Google+ είχε συγκεντρώσει 25 εκ. επισκέπτες και 540 εκ. ενεργούς χρήστες κάθε μήνα, τον Ιούνιο του 2014.

Στην πάροδο των επόμενων 2 χρόνων, τα smartphones και tablets άλλαξαν το πρόσωπο του social networking και τον τρόπο που επικοινωνούμε. Δεδομένου της ραγδαίας ανάπτυξης του mobile computing, δεν είναι περίεργο που τα περισσότερα social media εκμεταλλεύονται όλο και περισσότερο τις δυνατότητες των smartphone. Εφαρμογές όπως το Snapchat και το Instagram για διαμοίραση φωτογραφιών και βίντεο, δουλεύουν σχεδόν αποκλειστικά μόνο σε smartphones. Το ίδιο

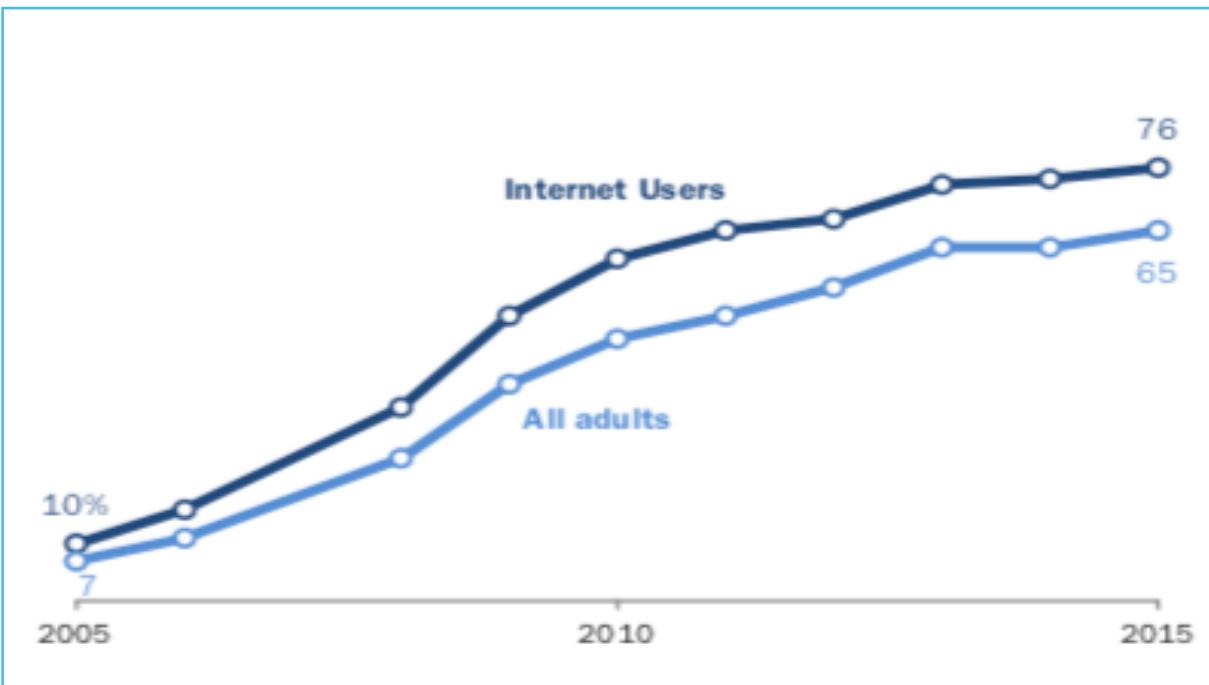
συμβαίνει με εφαρμογές σαν το Foursquare, μια εφαρμογή για την εύρεση τοποθεσιών και σημείων ενδιαφέροντος σε όλον το πλανήτη. Αντί να προσφέρουν μια γενική social networking εμπειρία, στοχεύουν και επικεντρώνονται σε ένα συγκεκριμένο είδος διάδρασης, όπως το μοίρασμα φωτογραφιών (Instagram), το ιδιωτικό μοίρασμα βίντεο (Snapchat), επαυξημένη πραγματικότητα (Foursquare) και εύρεση συντρόφων (Tinder).

Ουσιαστικά οι άνθρωποι χρησιμοποιούν έναν συνδυασμό από social media εφαρμογές για να χτίσουν μία εκτενή ψηφιακή ταυτότητα.

Τον Μάρτιο του 2014 το Facebook εξαγόρασε την Oculus VR, μια εταιρία μαζικής παραγωγής συσκευών εικονικής πραγματικότητας. Παρόλα αυτά το Facebook δεν έχει ασκήσει κάποια επιρροή στη διοίκηση της Oculus, αφήνοντας την εταιρία να ασχολείται κυρίως με το κομμάτι του gaming. Αυτή τη στιγμή ένα VR headset της Oculus κοστολογείται στα 300 εκ. \$. Η επιτυχία ή αποτυχία της Oculus στην αγορά είναι σίγουρο πως θα καθορίσει την στάση του Facebook αλλά και των άλλων μεγάλων κοινωνικών δικτύων προς τον τομέα της εικονικής πραγματικότητας. [13]

2.8 Στατιστικά Χρήσης των Κοινωνικών Δικτύων

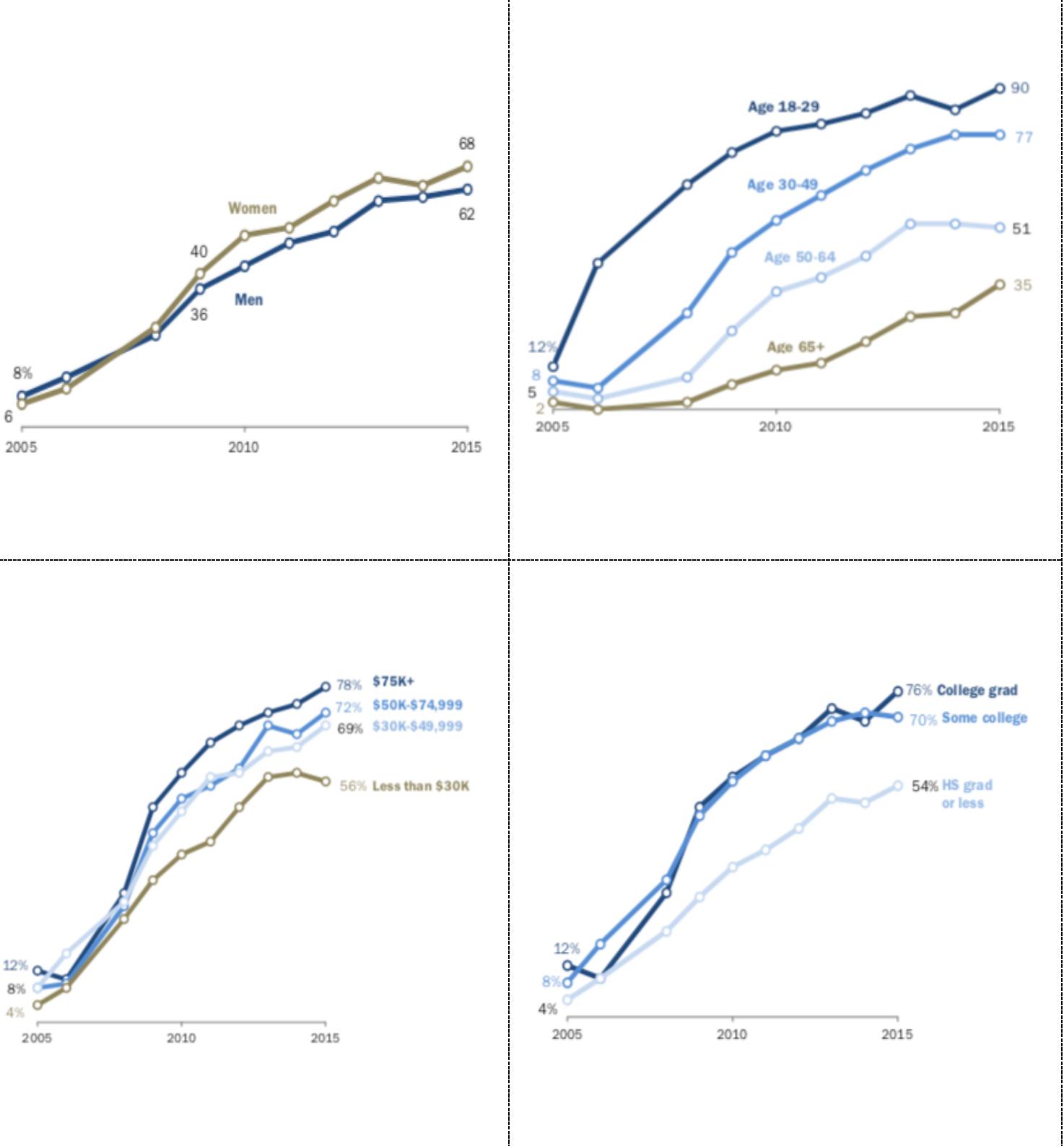
Σχεδόν το 65% των ενηλίκων στις ΗΠΑ χρησιμοποιεί sites κοινωνικής δικτύωσης, μια τάξη αύξησης 58% σε σχέση με το 7% που ήταν πριν 10 χρόνια. Αυτό είναι το κυριότερο συμπέρασμα της δημοσκόπησης του Pew Research Center για την χρήση των social media στο διάστημα 2005-2015. Παράλληλα μετρήθηκε πως η ανάπτυξη των social media επηρέασε τομείς όπως: η εργασία, η πολιτική καθώς και μοτίβα επικοινωνίας.



Εικόνα 2.4. % των Αμερικανών ενηλίκων και χρηστών Ίντερνετ

Μερικά ακόμα συμπεράσματα που προέκυψαν από την μελέτη είναι:

- **Ηλικιακές Διαφορές:** Οι νέοι ενήλικες (18-29) τείνουν να χρησιμοποιούν αρκετά τα κοινωνικά δίκτυα- σχεδόν το 90%. Όμως η χρήση σε άτομα ηλικίας 65 και άνω έχει τριπλασιαστεί από το 2010 που ήταν 11%. Το 2015 βρίσκονταν στο 35%, συγκριτικά με το 2% του 2005.
- **Διαφορές Φύλου:** Οι γυναίκες τείνουν να χρησιμοποιούν για περισσότερα χρόνια τα κοινωνικά δίκτυα σε σχέση με τους άνδρες, αν και η διαφορά είναι αρκετά μικρή. Το 68% των γυναικών τα χρησιμοποιεί σε σχέση με το 62% των ανδρών.
- **Κοινωνικο-οικονομικές Διαφορές:** Την περασμένη δεκαετία είχε παρατηρηθεί πως τα άτομα με υψηλότερα εισοδήματα είχαν την τάση να χρησιμοποιούν περισσότερο τα κοινωνικά δίκτυα. Πλέον όμως το 56% των ατόμων που ζει με χαμηλά αμειβόμενα εισοδήματα χρησιμοποιεί social media. Το ίδιο παρατηρείται και στο εκπαιδευτικό επίπεδο. Αυτοί που φοίτησαν σε κολλέγιο ή πανεπιστήμιο, ανεξάρτητα αν έλαβαν πτυχίο, είναι πιο πιθανό, να έχουν κάποιο λογαριασμό στα κοινωνικά δίκτυα σε σχέση με αυτούς που έχουν τελειώσει μόνο το λύκειο. [14]



Εικόνα 2.5. Γραφήματα κοινωνικών δικτύων σε σχέση με Φύλο-Ηλικία-Εισόδημα-Εκπαίδευση.

2.9 Social Media Marketing(SMM)

Η προβολή μέσω κοινωνικών δικτύων είναι πρακτική διαφήμισης που χρησιμοποιείται με σκοπό την απόκτηση δημοσιότητας και κίνησης μέσα από πηγές και ιστοσελίδες κοινωνικής δικτύωσης. Πρόκειται για μια πανίσχυρη στρατηγική, η οποία επιφέρει τεράστιους αριθμούς επισκεπτών, αναγνωρισμό της προϊόντος ή του λογοτύπου της εταιρείας, εύκολα και γρήγορα. Τα περισσότερα sites κοινωνικής δικτύωσης έχουν δικά τους εργαλεία ανάλυσης δεδομένων, τα οποία επιτρέπουν σε μία εταιρία να καταγράφει την πορεία και την επιτυχία των διαφημιστικών καμπανιών της. Σε στρατηγικό επίπεδο, το social media marketing εμπεριέχει τη διαχείριση της καμπάνιας, την διακυβέρνηση, τον ορισμό στρατηγικής (ενεργή ή παθητική προσέγγιση) και τον καθορισμό της “κουλτούρας” και “τόνου” της εταιρίας στα κοινωνικά δίκτυα.

Επίσης οι εταιρίες μέσω του SMM, επιτρέπουν στους χρήστες να ανεβάζουν περιεχόμενο δημιουργημένο από τους ίδιους (σχόλια, αξιολογήσεις προϊόντων, κ.α.), γνωστά και ως “earned media”, αντί να χρησιμοποιούν κλασσικές μεθόδους διαφήμισης. Οι εταιρίες χρησιμοποιούν αυτά τα άμεσα κανάλια επικοινωνίας ώστε να διαδρούν απευθείας με τους πελάτες τους. Αυτή η διάδραση είναι αρκετά πιο προσωπική σε σχέση με τις παραδοσιακές μεθόδους marketing και διαφήμισης. Η δυνατότητα που προσφέρει το ίντερνετ, μέσω ενός μηνύματος να επικοινωνείς με εκατομμύρια ανθρώπους, δημιουργησε μια e-word-of-mouth προοπτική, όπου οι πληροφορίες για ένα προϊόν επαναλαμβάνονται και διαδίδονται από πελάτη σε πελάτη. Όλο αυτό δημιουργεί περισσότερη “κίνηση” -traffic για την εταιρία ή το προϊόν.

Έτσι το SMM, το οποίο είναι πλέον βασικό κομμάτι του Internet Marketing, επικεντρώνεται στην εκμετάλλευση των μέσων κοινωνικής δικτύωσης (Social Media) με στόχο, τόσο την δημιουργία ενδιαφέροντος και μιας αμφίδρομης επικοινωνίας των χρηστών με το συγκεκριμένο brand/εταιρεία, όσο και με την υλοποίηση online διαφημιστικών / πρωθητικών ενεργειών και πλάνων. Έχοντας σαν βασικό εργαλείο τα κοινωνικά δίκτυα (Social Media), το SMM εστιάζει στην προσπάθεια δημιουργίας περιεχομένου που προσελκύει την προσοχή και ενθαρρύνει τους χρήστες να το μοιραστούν μέσω των κοινωνικών δικτύων τους.

Η χρήση των κινητών τηλεφώνων είναι αρκετά ωφέλιμη για το Social Media Marketing, επειδή δίνουν άμεση πρόσβαση στον παγκόσμιο ιστό και στα κοινωνικά δίκτυα. Πλέον έχει αλλάξει ριζικά το μονοπάτι προς την αγορά ενός προϊόντος, καθώς μέσω των κινητών, οι πελάτες ενημερώνονται σε πραγματικό χρόνο για τα προϊόντα και τις τιμές τους. Μέσω των QR codes μπορεί κανείς να έχει πρόσβαση στο site της εταιρίας ή σε online υπηρεσίες χρησιμοποιώντας ένα smartphone.

Επίσης, ο τρόπος με τον οποίο διαχωρίζονται / κατηγοριοποιούνται οι άνθρωποι αλλάζει ριζικά με την έλευση του κοινωνικού ιστού και των κοινωνικών μέσων δικτύωσης. Δημογραφικά στοιχειά όπως φύλο, ηλικία, εκπαίδευση και εισόδημα έχουν γίνει λιγότερο σημαντικά και αυτό που πραγματικά παίζει ρόλο είναι η κατηγοριοποίηση ανάλογα με τις συνήθειες και επιθυμίες των ανθρώπων, της συμπεριφοράς τους καθώς και των στάσεων και ενδιαφερόντων τους. [16] [15]

2.10 Στρατηγικές Social Media Marketing

Υπάρχουν 2 κύρια είδη στρατηγικών, η παθητική και η ενεργητική προσέγγιση, οι οποίες αναλύονται στη συνέχεια:

A) Παθητική Προσέγγιση

Τα κοινωνικά δίκτυα είναι μια αρκετά χρήσιμη πηγή πληροφοριών της αγοράς και ένας τρόπος να αντιληφθεί μια εταιρία την ματιά και τον τρόπο σκέψης των πελατών της. Τα Blogs, οι κοινότητες και φόρουμς είναι πλατφόρμες όπου κανείς μπορεί να μοιραστεί την γνώμη του και τις κριτικές του για εταιρίες, προϊόντα και υπηρεσίες. Οι εταιρίες έτσι μπορούν να αναλύσουν την φωνή των πελατών τους και να λάβουν άμεσο feedback από τα κοινωνικά δίκτυα για διαφημιστικούς σκοπούς. Το σημαντικότερο χαρακτηριστικό, είναι πως το social media marketing είναι μια γενικά οικονομική μορφή ανάλυσης της αγοράς και του target group μιας εταιρίας που οδηγεί σε γρήγορη αναγνώριση προβλημάτων αλλά και ευκαιριών.

Για παράδειγμα, η Apple μετά την κυκλοφορία του iPhone 6, είχε αρκετές κακές κριτικές και παράπονα από χρήστες που ισχυριζόντουσαν πως το νέο iPhone λυγίζει με λίγη πίεση. Η Apple αμέσως εξέδωσε ανακοινώσεις στα κοινωνικά δίκτυα λέγοντας πως είναι εξαιρετικά σπάνιο φαινόμενο και πως θα κινηθεί κατάλληλα για την επισκευή του προβλήματος. Αντίθετα, σε μια παραδοσιακή marketing προσέγγιση θα έπρεπε να γίνει έρευνα, να καθοριστεί η “ομάδα-στόχος” και αρκετή εξόρυξη πληροφοριών και δεδομένων. Όλο αυτό εκτός ότι θα διαρκούσε αρκετές εβδομάδες μέχρι η εταιρία να αντιληφθεί το πρόβλημα, την συμπεριφορά των καταναλωτών και τη γνώμη αυτών προς την εταιρία θα ήταν επίσης και εξαιρετικά δαπανηρό.

B) Ενεργητική Προσέγγιση

Τα social media μπορούν να χρησιμοποιηθούν ως δίαυλος κοινωνικών σχέσεων-εργαλεία marketing αλλά και ως κανάλια άμεσης επικοινωνίας στοχευμένα σε συγκεκριμένο αγοραστικό κοινό. Χρησιμοποιώντας εργαλεία όπως το Google Adsense, οι διαφημιστές μπορούν να στοχεύσουν τις διαφημίσεις τους σε ακόμα πιο συγκεκριμένο κοινό, όπως π.χ. άτομα που ενδιαφέρονται για entrepreneurship, πολιτική ή είναι συσχετισμένοι με κάποιο πολιτικό κόμμα ή ιδεολογία. Αυτό το

κάνει το AdSense μέσω της αναζήτησης και χρήσης των online post ή των σχολίων που κάνει κάποιος στα κοινωνικά δίκτυα

Τα κοινωνικά δίκτυα μπορούν ουσιαστικά να χρησιμοποιηθούν ως εργαλείο για φτηνές έρευνες αγοράς, καθώς μια εταιρία μπορεί να τα χρησιμοποιήσει εύκολα και γρήγορα για να επικοινωνήσει το προϊόν της σε πολλά άτομα, μέσω ενός ενδιάμεσου διάσημου ατόμου. Για παράδειγμα ένας αθλητής ο οποίος σπονσοράρει μια εταιρία, αυτόματα διαθέτει τους οπαδούς του ως εν δυνάμη αγοραστικό κοινό για την εταιρία, καθώς πολλοί οπαδοί θα θέλουν να μιμηθούν το lifestyle του αθλητή ή και να φτάσουν το επίπεδο του αθλητή χρησιμοποιώντας το προϊόν αυτό. [15]

Σύνοψη Κεφαλαίου

Στη σημερινή κοινωνία η έμφυτη ανάγκη των ανθρώπων για επικοινωνία έχει κάνει τα κοινωνικά δίκτυα αρκετά δημοφιλή και αναπόσπαστο κομμάτι της ζωής μας. Μέσα από τα κοινωνικά δίκτυα μπορεί κανείς να δημιουργήσει ένα δημόσιο προφίλ και να επικοινωνεί με ένα πλήθος από άλλους χρήστες με τους οποίους μοιράζεται την ίδια μορφή σύνδεσης. Ένα κοινωνικό δίκτυο χαρακτηρίζεται από 5 βασικά στοιχεία: 1) Είναι βασισμένο πάνω στους χρήστες, 2) υπάρχει αλληλεπίδραση μεταξύ των χρηστών, 3) οι οποίοι δημιουργούν κοινότητες, 4) σχέσεις, 5) και εκφράζουν τα συναισθήματα τους. Τα κοινωνικά δίκτυα έχουν αρκετές χρησιμότητες. Εξυπηρετούν την σύνδεση με υπάρχοντα δίκτυα και φιλίες, διαδικτυακή αντιπροσώπευση, βοηθούν στην εύρεση πληροφοριών καθώς και στη δημιουργία περιεχομένου. Μερικά από τα δημοφιλέστερα κοινωνικά δίκτυα αυτή τη στιγμή είναι το Facebook, Twitter, LinkedIn και Instagram με περισσότερους από 2,5 δισεκατομμύρια χρήστες συνολικά. Φυσικό επακόλουθο είναι τα κοινωνικά δίκτυα να έχουν επηρεάσει τομείς όπως: η επικοινωνία, η εργασία, η πολιτική αλλά και η διαφήμιση. Μέσω του social media marketing μπορούν οι εταιρίες να οργανώσουν καλύτερα τις διαφημιστικές καμπάνιες τους και να απευθυνθούν άμεσα στους πελάτες τους.

Κεφάλαιο 3

Τα Δεδομένα και η Επεξεργασία τους

3.1 Εισαγωγή

Ως εποχή της πληροφορίας ή αλλιώς ψηφιακή εποχή χαρακτηρίζεται η μετάβαση από την παραδοσιακή βιομηχανία που είχε ξεκινήσει με την Βιομηχανική Επανάσταση σε μία οικονομία που βασίζεται, ως επί το πλείστον, σε τεχνολογίες της πληροφορίας. Η βάση της ανωτέρω επανάστασης βασίζεται στην ψηφιοποίηση της πληροφορίας. Η τεχνολογική πρόοδος επιτρέπει την αποθήκευση μεγάλου όγκου πληροφοριών σε ηλεκτρονική μορφή. Ο μεγάλος όγκος των πληροφοριών μπορεί πλέον να μετατραπεί σε δεδομένο.

Ως δεδομένα μπορεί να ορίσει κανείς τα σημεία πληροφοριών επί επιστημονικών παρατηρήσεων ή συμπεριφορών και να περιλαμβάνουν λέξεις - έννοιες, αριθμούς, σύμβολα, διαγράμματα, σχέδια, φωτογραφίες, κλπ. που περιγράφουν ή αντιπροσωπεύουν ποσότητες, έννοιες, ιδέες, αντικείμενα, γεγονότα, καταστάσεις και λειτουργίες. Ενδεχομένως κάποιοι από τους τύπους δεδομένων που παρατίθενται εμπεριέχουν ήδη εμφανείς πληροφορίες, όχι όμως την πληροφορία στο επίπεδο που συνθέτει μια αξιολόγησή τους. Με την προσθήκη αξιών στα ηλεκτρονικά δεδομένα, δηλαδή με διαμόρφωση, οργάνωση, μαθηματική ή στατιστική ανάλυση, διόρθωση λαθών ή συμπίεση παράγεται η ηλεκτρονική πληροφορία. Γενικότερα, δεδομένο ονομάζεται ένα γνωστό ή αποδεκτό στοιχείο το οποίο χρησιμοποιείται ως βάση ή προϋπόθεση στην επίλυση προβλημάτων.

[17]

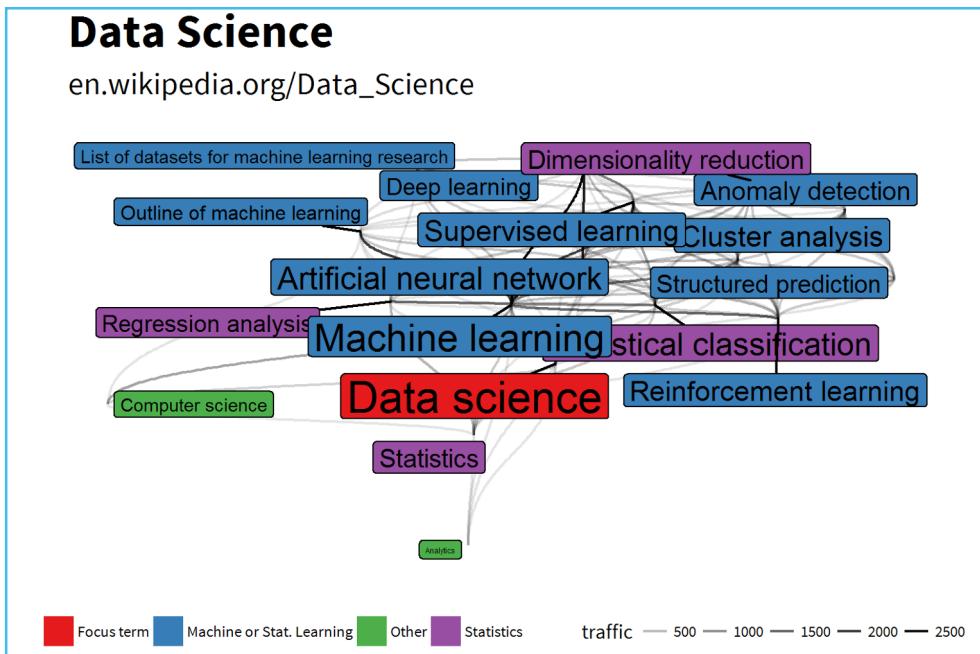
Ενώ η έννοια των δεδομένων πολλές φορές συγχέεται με την επιστημονική έρευνα, τα δεδομένα συγκεντρώνονται από ένα τεράστιο εύρος οργανισμών, κυβερνήσεων και εταιριών (αριθμός πωλήσεων, τιμές μετοχών, εισοδήματα, ποσοστά ανεργίας, συχνότητα καιρικών φαινομένων). Τα δεδομένα μετριούνται, συγκεντρώνονται, αναφέρονται και αναλύονται. Τις περισσότερες φορές οπτικοποιούνται με χρήση γράφων, διαγραμμάτων, εικόνων και άλλων εργαλείων. Ως ευρύτερη έννοια αναφέρονται στο γεγονός πως υφίσταται κάποια πληροφορία-γνώση, η οποία μπορεί να μετατραπεί σε μία καλύτερη μορφή πληροφορίας για την εξαγωγή ενός συμπεράσματος.

Τα “Raw Data” ή αλλιώς ανεπεξέργαστα δεδομένα είναι μία συλλογή αριθμών ή χαρακτήρων που πρέπει να “καθαριστεί”, να ταξινομηθεί και να δομηθεί από τους ερευνητές ώστε να μπορούν να αναλυθούν με κάποιο μοντέλο. Πριν την ανάπτυξη των υπολογιστών, τα δεδομένα συγκεντρώνονταν κυρίως από αναλυτές. Από το 2010 και μετά χρησιμοποιούνται κατά κύριο λόγο

υπολογιστές για τη συγκέντρωση δεδομένων, ταξινόμηση και την επεξεργασία αυτών, σε τομείς όπως πχ. το μάρκετινγκ ή ανάλυση χρήσης κοινωνικών υπηρεσιών. Τα μοτίβα και τα συμπεράσματα που εξάγονται από τα δεδομένα μπορούν να χρησιμοποιηθούν ως πληροφορία για τη βελτίωση της υπάρχουσας γνώσης σε ένα πεδίο έρευνας.[18]

3.1.1 Η Επιστήμη των Δεδομένων

Η επιστήμη των δεδομένων ή αλλιώς Data Science είναι το διεπιστημονικό πεδίο που χρησιμοποιεί μεθόδους, αλγόριθμους και συστήματα για να εξάγει γνώση και πληροφορίες από τα δεδομένα είτε σε δομημένη ή αδόμητη μορφή. Η επιστήμη των δεδομένων ενοποιεί τη στατιστική με την ανάλυση δεδομένων και τη μηχανική μάθηση ώστε ο επιστήμονας να κατανοήσει και να αναλύσει πραγματικά φαινόμενα με δεδομένα. Οι τεχνικές και οι μέθοδοι που χρησιμοποιούνται, αντλούνται από πεδία όπως: των μαθηματικών, της στατιστικής και της πληροφορικής.



Εικόνα 3.1. Η συσχέτιση της επιστήμης των δεδομένων με άλλα πεδία.

Η επιστήμη αυτή είναι σχετικά πρόσφατη αν σκεφτεί κανείς πως ο όρος data science δημιουργήθηκε στα μέσα της δεκαετίας 2000-2010. Το 2012 το Harvard Business Review χαρακτήρισε την επιστήμη των δεδομένων ως “την πιο ελκυστική δουλειά του 21ου αιώνα” και από τότε η απήχηση του όρου αυξάνεται συνεχώς. Ο όρος data science κατά κάποιο τρόπο κάλυψε και εσώκλεισε τους όρους “ανάλυση επιχειρήσεων”, “μοντέλα πρόβλεψης” και “στατιστική”.

3.1.2 Big Data

Τα “μεγάλα δεδομένα” ή αλλιώς Big Data είναι συλλογές δεδομένων τόσο μεγάλες και περίπλοκες που τα παραδοσιακά συστήματα ανάλυσης δεδομένων αποτυγχάνουν να τα χειριστούν στα πλαίσια ενός εύλογου χρονικού πλαισίου. Στα Big Data συμπεριλαμβάνονται τόσο αδόμητες, ημιδομημένες όσο και δομημένες συλλογές δεδομένων, αν και τις περισσότερες φορές πρόκειται για χειρισμό αδόμητων δεδομένων. Το μέγεθος των Big Data είναι ένας όλο και αυξανόμενος αριθμός, που το 2012 κυμαινόταν στις μερικές δεκάδες terabytes, πλέον ο αριθμός αυτός φτάνει τα πολλά exabytes. Τα μεγάλα δεδομένα απαιτούν μια γκάμα εργαλείων και τεχνολογιών που μέσα από σύνθετα, ποικίλα και ογκώδη δεδομένα αποκαλύπτουν χρήσιμα στοιχεία.

Τα χαρακτηριστικά τα οποία μπορούν να περιγράψουν τα Big Data είναι τα εξής :

- **Όγκος(Volume):** Το μέγεθος των δεδομένων καθορίζει την ποιότητα των εξαγόμενων πληροφοριών και κατά πόσο η αρχική συλλογή χαρακτηρίζεται ότι μη Big Data.
- **Ποικιλία(Variety):** Ο τύπος και η φύση των δεδομένων. Αυτό βοηθά στην αποδοτική ανάλυση των εξαγόμενων πληροφοριών από τα αρχικά κείμενα, εικόνες, βίντεο.
- **Ταχύτητα(Velocity):** Η ταχύτητα με την οποία παράγονται τα δεδομένα και επεξεργάζονται για να συμβαδίζουν με τις απαιτήσεις και τις προκλήσεις.
- **Εγκυρότητα(Validity):** Η ποιότητα των αρχικών δεδομένων μπορεί να ποικίλει, επηρεάζοντας έτσι την σωστή ανάλυση.

Το 2004 η Google δημοσίευσε μια έρευνα σχετικά με τον αλγόριθμο MapReduce. Το MapReduce παρείχε ένα μοντέλο παράλληλης επεξεργασίας και η υλοποίησή του έγινε για την επεξεργασία τεράστιου όγκου δεδομένων. Με αυτήν την τεχνική τα δεδομένα χωρίζονται και μοιράζονται σε παράλληλους κόμβους επεξεργασίας. Στη συνέχεια τα αποτελέσματα συγκεντρώνονται και μοιράζονται πάλι πίσω. Η δομή αυτή ήταν τόσο επιτυχής που και άλλες εταιρίες θέλησαν να την επαναλάβουν όπως για παράδειγμα η Apache που χρησιμοποιώντας αυτό το framework δημιούργησε το Hadoop. Μελέτες του 2012 δείχνουν ότι μία αρχιτεκτονική πολλαπλών στρωμάτων έχει εξίσου καλά αποτελέσματα όπως μια αρχιτεκτονική παράλληλης διαμοίρασης.

Τα τελευταία χρόνια έχει αυξηθεί η ζήτηση για σωστό χειρισμό δεδομένων τόσο που οι μεγάλες εταιρίες τεχνολογίας όπως IBM, HP, Microsoft κ.α. έχουν ξοδέψει πάνω από 15 δις δολάρια σε εφαρμογές που χειρίζονται και αναλύουν δεδομένα και πληροφορίες. Η “βιομηχανία των δεδομένων” εκτιμόταν πως άξιζε 100 δις δολάρια το 2010 και κάθε χρόνο αυξανόταν 10%, ξεπερνώντας κατά πολύ όλη τον τομέα του “software business”. Ανάμεσα από το 1990 και 2005 εκτιμάται πως 1 δις άνθρωποι μετακινήθηκαν στην μεσαία τάξη, που σημαίνει μεγαλύτερος αριθμός

ανθρώπων είχε πρόσβαση στην εκπαίδευση και ως αποτέλεσμα αύξηση του πλήθος δεδομένων. Σχεδόν το ένα τρίτο της παγκόσμιας αποθηκευμένης πληροφορίας είναι σε αλφαριθμητικό κείμενο ή εικόνα. Αυτή είναι και η πιο χρήσιμη μορφή για τις big data εφαρμογές. Αυτό επίσης δείχνει και την προοπτική των μη μαζικά χρησιμοποιούμενων ακόμα δεδομένων (βίντεο και ήχος). [19]

Τα μεγάλα δεδομένα έχουν μια ευρεία γκάμα εφαρμογών και χρησιμοποιούνται πλέον σε όλους τους τομείς. Μερικές αναφορές εφαρμογών είναι οι εξής:

- **Κυβερνήσεις:** Η χρήση και η υιοθέτηση των μεγάλων δεδομένων έχει αρκετά κυβερνητικά οφέλη. Οι πληροφορίες που εξάγονται βοηθούν στη λήψη αποφάσεων για μείωση του κόστους, αύξηση παραγωγικότητας και καινοτομίας.
- **Διεθνής ανάπτυξη:** Βελτιώσεις στην ανάλυση των μεγάλων δεδομένων προσφέρουν χαμηλού κόστους ευκαιρίες για να βελτιωθεί η λήψη αποφάσεων σε θέματα υγείας, ανεργίας, εγκληματικότητας και φυσικών καταστροφών.
- **Βιομηχανία:** Η ανάλυση των μεγάλων δεδομένων παρέχει βελτιώσεις στον σχεδιασμό της γραμμής παραγωγής και της γραμμής τροφοδοσίας.
- **Υγεία:** Ο τομέας της υγείας με την χρήση των Big Data βελτιώνεται συνεχώς, καθώς μπορούν να προβλέψουν αύξηση ασθενειών, να κάνουν καλύτερη διαχείριση των ασθενών και να μειώσουν τα λειτουργικά κόστη.
- **Media:** Στελέχη στα μέσα μαζικής ενημέρωσης και στη διαφήμιση χειρίζονται τα μεγάλα δεδομένα ως σημεία πρόσβασης για εκατομμύρια κόσμου. Η βιομηχανία της διαφήμισης απομακρύνεται πλέον από τις κλασικές μεθόδους προσέγγισης όπως: τηλεόραση, εφημερίδες, περιοδικά και επενδύει σε τεχνολογίες που στοχεύουν στο αγοραστικό κοινό την κατάλληλη χρονική στιγμή και τοποθεσία.
- **Ασφάλεια:** Οι ασφαλιστικές εταιρίες συγκεντρώνουν δεδομένα από στοιχεία και κριτήρια που καθορίζουν την “κατάσταση υγείας” ενός ατόμου, όπως: διατροφικές συνήθειες, οικογενειακή κατάσταση και αγοραστικές συνήθειες, από τις οποίες γίνονται προβλέψεις για το κόστος περίθαλψης των ασφαλισμένων.
- **Internet of Things (IoT):** Τα δεδομένα που εξάγονται από τις συσκευές IoT παρέχουν συλλογές που χρησιμοποιούνται από τη βιομηχανία μαζικής ενημέρωσης αλλά και κυβερνήσεις ώστε να αυξήσουν την ακροαματικότητά τους και την αποδοτικότητά τους.

[20]

3.1.3 GDPR Κανονισμός Προστασίας Δεδομένων

Τα δεδομένα έχουν, και όχι άδικα, χαρακτηριστεί “το νόμισμα του 21ου αιώνα”. Ως δεδομένο στην ψηφιακή εποχή μπορεί να οριστεί κάθε πληροφορία η οποία έχει διαμορφωθεί με τέτοιον τρόπο, ώστε να είναι αποτελεσματική η μεταφορά και η επεξεργασία της. Προσωπικό μπορεί να είναι κάθε δεδομένο. Η φύση της περιεχόμενης πληροφορίας μπορεί να αφορά καθετί χαρακτηρίζει ένα άτομο: καλύπτει τόσο αντικειμενικές (ύψος, χρώμα κτλ.) όσο και υποκειμενικές (απόψεις και δηλώσεις) πληροφορίες. Δεν απαιτείται οι πληροφορίες να είναι εξακριβωμένες ή αληθείς. Όσον αφορά το μέσο που παρέχεται η πληροφορία, δεδομένα κάθε μορφής καλύπτονται από τον ορισμό πχ. αλφαριθμητικά, γραφικά, ακουστικά δεδομένα, έγχαρτα ή σε ηλεκτρονική μορφή. Τα προσωπικά δεδομένα που συνδέονται με τα θεμελιώδη δικαιώματα και τις ελευθερίες των υποκειμένων χρήζουν ειδικής, αυξημένης προστασίας και καλούνται ευαίσθητα προσωπικά δεδομένα. Η επεξεργασία αυτών επιτρέπεται μόνο για πολύ συγκεκριμένους σκοπούς και υπό την παροχή επαρκών εγγυήσεων.

Με μια διευρυμένη ερμηνεία, ένα δεδομένο είναι προσωπικό, αν αφορά την ταυτότητα, τα χαρακτηριστικά ή την συμπεριφορά του ατόμου ή αν οι πληροφορίες αυτές χρησιμοποιούνται για να διαπιστωθεί ή να επηρεαστεί ο τρόπος που το άτομο αυτό χαρακτηρίζεται ή αντιμετωπίζεται. Μόνο ζώντα πρόσωπα δύναται να διαθέτουν προσωπικά δεδομένα, καθώς τα στοιχεία των νεκρών προσώπων αποτελούν στατιστικά, ιστορικά ή δεδομένα συναισθηματικής φύσεως.

Η διαδικασία της επεξεργασίας έχει διάφορα στάδια, τα οποία μπορεί να περιλαμβάνουν: επαλήθευση, τακτοποίηση με κάποια σειρά, κατηγοριοποίηση, περίληψη, αναφορά, ανάλυση και ερμηνεία των δεδομένων.

Για να γίνει σαφής η ανάγκη της προστασίας των προσωπικών δεδομένων, πρέπει να διασαφηνιστεί η σχέση της ιδιωτικότητας με τα προσωπικά δεδομένα. Η ιδιωτικότητα είναι η δικαιολογητική βάση για την απαγόρευση της κρατικής παρέμβασης στον ιδιωτικό και οικογενειακό βίο και εμφανίζεται σαν θεμέλιο για την ελευθερία της σκέψης και τις προσωπικές επιλογές του ατόμου. Η ιδιωτικότητα εκτείνεται σε τόσο πολυποίκιλα ζητήματα και επομένως είναι αναγκαίος ο χαρακτηρισμός της ως μια ιδέα προσωπικού χώρου, μέσα στον οποίο κάθε άτομο μπορεί ελεύθερα και αυτόνομα να κάνει τις επιλογές του.

Ο κανονισμός (ΕΕ) 2016/679 του Ευρωπαϊκού Κοινοβουλίου και του Συμβουλίου για την προστασία των φυσικών προσώπων έναντι της επεξεργασίας των δεδομένων προσωπικού χαρακτήρα και για την ελεύθερη κυκλοφορία των δεδομένων αυτών αποτελεί το θεσμικό εργαλείο που ρυθμίζει την προστασία των δεδομένων. Στόχος του κανονισμού είναι η θέσπιση ενός υψηλού επιπέδου προστασίας δεδομένων χωρίς να θίγει την ελεύθερη κυκλοφορία δεδομένων εντός της Ε.Ε.

Μέσω του κανονισμού διασφαλίζεται η συνεκτική και η ομοιόμορφη εφαρμογή των σχετικών κανόνων για την προστασία έναντι της επεξεργασίας των δεδομένων προσωπικού χαρακτήρα σε ολόκληρη την ΕΕ.

Η εφαρμογή του κανονισμού δεν περιορίζεται εντός των ορίων της ΕΕ, καθώς δεν αγνοείται η διεθνοποίηση της οικονομίας, η ύπαρξη συμφερόντων που εκτείνονται εκτός των φυσικών συνόρων της Ευρώπης και η ανάγκη για την διασυνοριακή ροή πληροφοριών χωρίς να κάμπτεται το επιθυμητό επίπεδο προστασίας τους. Ο Κανονισμός εφαρμόζεται επομένως: 1) σε κάθε υπεύθυνο επεξεργασίας ή εκτελούντα που διενεργεί επεξεργασία δεδομένων στο πλαίσιο δραστηριοτήτων του και έχει εγκατάσταση εντός της ΕΕ, και 2) σε κάθε υπεύθυνο επεξεργασίας ή εκτελούντα ανεξαρτήτως έδρας εφόσον επεξεργάζεται δεδομένα υποκείμενων που διαμένουν στην ΕΕ.

Η νομιμότητα επεξεργασίας δεδομένων μπορεί να θεμελιώνεται ως εξής:

1. **Συγκατάθεση:** Η επεξεργασία είναι νόμιμη, εφόσον το υποκείμενο έχει συναινέσει στην επεξεργασία των δεδομένων του για συγκεκριμένους σκοπούς. Πιο συγκεκριμένα οι προϋποθέσεις εγκυρότητας της συγκατάθεσης είναι:
 - **Ελεύθερη:** Η συγκατάθεση θα πρέπει να είναι προϊόν ελεύθερης επιλογής του υποκειμένου, το οποίο έχει δικαίωμα να αρνηθεί να παράσχει την συγκατάθεσή του.
 - **Εν επιγνώσει:** Η συγκατάθεση πρέπει να είναι ενημερωμένη, γεγονός που απαιτεί να παρέχονται στο υποκείμενο με κατανοητό τρόπο και με ακρίβεια όλες οι αναγκαίες πληροφορίες που σχετίζονται με την επεξεργασία.
 - **Συγκεκριμένη:** Η συγκατάθεση πρέπει να δίνεται για συγκεκριμένο σκοπό επεξεργασίας. Αν οι σκοποί επεξεργασίας είναι περισσότεροι, η συγκατάθεση πρέπει να δοθεί ξεχωριστά για τον καθένα.
 - **Δήλωση ή σαφής θετική ενέργεια:** Η συγκατάθεση πρέπει να δίνεται με δήλωση ή τουλάχιστον με σαφή θετική ενέργεια που δεν αφήνει περιθώρια αμφιβολίας.
2. Η επεξεργασία είναι απαραίτητη για την εκτέλεση της σύμβασης ή λήψης μέτρων κατά το προσυμβατικό στάδιο κατ' αίτηση του υποκειμένου.
3. Η επεξεργασία είναι απαραίτητη για σκοπούς έννομων συμφερόντων του υπευθύνου επεξεργασίας ή τρίτου, εφόσον κατόπιν στάθμισης προκύπτει ότι αυτά υπερισχύουν έναντι των δικαιωμάτων του υποκειμένου.
4. Η επεξεργασία είναι απαραίτητη για τη συμμόρφωση με έννομη υποχρέωση που υπέχει ο υπεύθυνος επεξεργασίας βάση του ενωσιακού ή εθνικού δικαίου και όχι βάσει σύμβασης.

5. Η επεξεργασία είναι απαραίτητη για την εκπλήρωση καθήκοντος προς το δημόσιο συμφέρον ή κατά την άσκηση δημόσιας εξουσίας που έχει ανατεθεί στον υπεύθυνο.
6. Η επεξεργασία είναι απαραίτητη για τη διαφύλαξη ζωτικού συμφέροντος του υποκειμένου των δεδομένων ή άλλου φυσικού προσώπου. [21] [22]

3.2 Εξόρυξη Δεδομένων (Data Mining)

Η εξόρυξη δεδομένων είναι ένα πολυσυζητημένο θέμα στις μέρες μας. Σε αντίθεση με τα προηγούμενα χρόνια, πλέον όλα είναι συνδεδεμένα με τα δεδομένα και είμαστε ικανοί να χειριστούμε τα δεδομένα. Συλλέγοντας και παρατηρώντας αυτά τα σύνολα δεδομένων, οι αναλυτές μπορούν να εντοπίσουν μοτίβα. Ακόμα και αν όλο το σετ δεδομένων χαρακτηρίζεται ως “σκουπίδι” (junk), υπάρχουν πάντα κάποια κρυφά πρότυπα που μπορούν να εξαχθούν συνδυάζοντας πολλές πηγές δεδομένων. [23]

Πιο συγκεκριμένα, εξόρυξη δεδομένων είναι η διαδικασία εύρεσης προτύπων σε μεγάλες συλλογές δεδομένων συμπεριλαμβανομένου μεθόδων και εφαρμογών της μηχανικής μάθησης, στατιστικής, και συστημάτων βάσεων δεδομένων. Η εξόρυξη δεδομένων είναι το διεπιστημονικό πεδίο της επιστήμης των υπολογιστών με γενικότερο στόχο την εξαγωγή πληροφορίας από τα αρχικά δεδομένα και την μετατροπή αυτών σε γνώση με περιεκτική δομή για μετέπειτα χρήση. Πέρα από το βήμα της ανάλυσης, στην όλη διαδικασία εμπεριέχονται βάσεις δεδομένων και ο χειρισμός τους, προ-επεξεργασία δεδομένων, θεωρήσεις συμπερασμάτων και θεωριών, μετρήσεις, μετεπεξεργασία δεδομένων και τέλος η οπτικοποίηση (visualization).

Το πραγματικό έργο της Εξόρυξης Δεδομένων είναι η αυτοματοποιημένη ανάλυση μεγάλων ποσοτήτων δεδομένων για την εξαγωγή μέχρι πρότινος άγνωστων, ενδιαφέροντων προτύπων όπως μία ομάδα δεδομένων (cluster analysis), ασυνήθιστων δεδομένων (anomaly detection) και συσχετίσεων ανάμεσα στα δεδομένα.

Η διαδικασία της εξόρυξης δεδομένων περιλαμβάνει τα βήματα :

1. **Συλλογή δεδομένων:** Πριν γίνει χρήση των αλγορίθμων εξόρυξης δεδομένων πρέπει να καθοριστεί το κατάλληλο σύνολο, το οποίο ταυτόχρονα να είναι αρκετά μεγάλο ώστε να περιέχει τα ζητούμενα πρότυπα αλλά και να παραμένει συνοπτικό για να εξορυχθεί σε αποδεκτό χρονικό διάστημα.
2. **Καθαρισμό των δεδομένων:** Από τη στιγμή που αναφερόμαστε σε μεγάλα δεδομένα, πρέπει να εξασφαλιστεί ότι περιέχονται τα απαραίτητα δεδομένα και έχουν αφαιρεθεί τα ανεπιθύμητα. Η προ-επεξεργασία του καθαρισμού είναι κομβικό σημείο για την

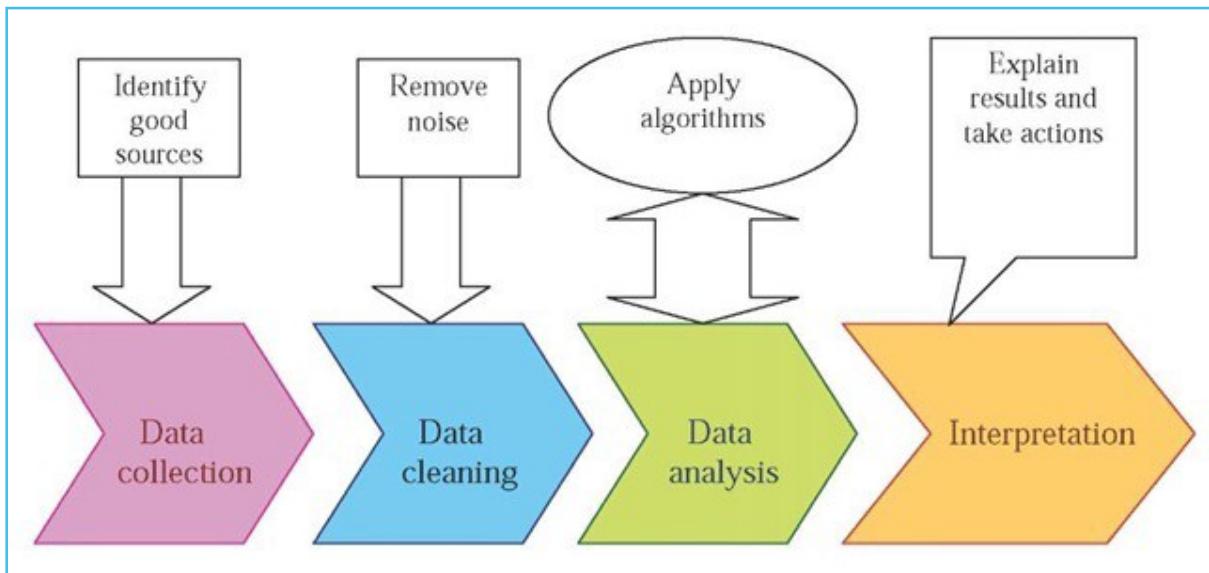
ανάλυση δεδομένων με πολλούς παράγοντες, καθώς διαγράφονται στοιχεία που περιέχουν “θόρυβο” και ελλιπείς πληροφορίες.

3. Εξόρυξη και Ανάλυση των δεδομένων: Σε αυτό το βήμα περιέχονται οι εξής 6 ενέργειες:

- **Ανίχνευση Ανωμαλιών:** Η ταυτοποίηση ασυνήθιστων δεδομένων, τα οποία περιέχουν σφάλματα και ανακρίβειες. Επίσης μπορεί τα δεδομένα να περιέχουν ενδιαφέρουσες τιμές ώστε να χρειάζονται περαιτέρω εξακρίβωση.
- **Κανόνας συνδυαστικής μάθησης (dependency modelling):** Γίνεται αναζήτηση για συσχετίσεις μεταξύ των δεδομένων. Π.χ. ένα σουπερμάρκετ συγκεντρώνει δεδομένα για τις αγοραστικές συνήθειες των πελατών του. Χρησιμοποιώντας αυτόν τον κανόνα μπορεί να διακρίνει ποια προϊόντα αγοράζονται συχνά μαζί και να προβεί σε αντίστοιχες προωθητικές ενέργειες.
- **Ομαδοποίηση (Clustering):** Ανακαλύπτονται ομάδες και δομές στα δεδομένα, οι οποίες κατά κάποιο τρόπο εμφανίζουν κοινά χαρακτηριστικά, χωρίς να χρησιμοποιούνται ήδη γνωστές δομές στα δεδομένα.
- **Ταξινόμηση (Classification):** Εδώ κατατάσσονται τα δεδομένα σε γενικές κατηγορίες. Π.χ. τα E-mails κατατάσσονται σε “αποδεκτά” ή “σπαμ”.
- **Αναδρομή:** Γίνεται αναζήτηση μιας μεθόδου που μοντελοποιεί τα δεδομένα με το ελάχιστο δυνατό σφάλμα για την εκτίμηση των σχέσεων ανάμεσα στα δεδομένα.
- **Σύνοψη:** Εμπεριέχει την οπτικοποίηση (visualization) και μια συμπαγή αναπαράσταση των δεδομένων.

4. Ερμήνευση -Επαλήθευση δεδομένων: Το τελευταίο βήμα είναι η εξακρίβωση των προτύπων που εξήχθησαν από τον αλγόριθμο εξόρυξης σε ένα ευρύτερο σύνολο δεδομένων. Δεν είναι αναγκαίο όλα τα πρότυπα και οι συσχετίσεις που θα εξαχθούν να είναι έγκυρα. Για να ξεπεραστεί αυτό το πρόβλημα, εκπαιδεύεται σε ένα δοκιμαστικό σύνολο δεδομένων αρχικά. Π.χ. για έναν αλγόριθμο εύρεσης spam (ενοχλητικών) emails θα του δοθεί πρώτα ένα σύνολο από δοκιμαστικά emails. Έπειτα αφού εκπαιδευτεί ο

αλγόριθμος θα του δοθεί ένα νέο άγνωστο σύνολο από emails για να μετρηθεί η ακρίβεια εύρεσης των προτύπων. [24]



Εικόνα 3.2. Τα Βασικά βήματα της εξόρυξης δεδομένων

3.2.1 Εξόρυξη Δεδομένων στα Κοινωνικά Δίκτυα

Η εξόρυξη δεδομένων στα κοινωνικά δίκτυα ή αλλιώς social media mining είναι η διαδικασία συλλογής μεγάλων δεδομένων, παραγμένων από χρήστες κοινωνικών δικτύων ή εφαρμογών σε smartphones με σκοπό την εξαγωγή μοτίβων και συμπερασμάτων σχετικά με τους χρήστες και τη χρήση αυτής της πληροφορίας για διαφημιστικούς ή ερευνητικούς σκοπούς. Στο social media mining ο ρόλος του αναλυτή δεδομένων και των εφαρμογών ανάλυσης είναι ιδιαίτερα σημαντικός, καθώς γίνεται συλλογή τεράστιου όγκου αδόμητης πληροφορίας, από την οποία θα προκύψουν σχέσεις μεταξύ χρηστών, τάσεις και μοτίβα, συμπεριφορές, αγοραστικές συνήθειες και πολλά άλλα. Αυτές οι τάσεις και τα μοτίβα αφορούν κυρίως εταιρίες, οργανισμούς και μη κερδοσκοπικά ιδρύματα με σκοπό να βελτιώσουν τις πολιτικές προσέγγισης τους αλλά και για να διεξάγουν έρευνες.

Η εξόρυξη δεδομένων στα κοινωνικά δίκτυα είναι ένα πεδίο εφαρμογής της εξόρυξης δεδομένων, που χρησιμοποιεί βασικές έννοιες από την επιστήμη των υπολογιστών, τη μηχανική μάθηση και τη στατιστική. Οι αναλυτές-εκσκαφείς (miners) αναπτύσσουν αλγορίθμους κατάλληλους για να χειρίζονται μεγάλα αρχεία που περιέχουν σχόλια, φωτογραφίες, αντιδράσεις, τοποθεσίες, ηλικίες και πολλά άλλα στοιχεία χρηστών από τα κοινωνικά μέσα δικτύωσης. Για την ανάλυση όλων αυτών των στοιχείων χρησιμοποιούνται μέθοδοι και κανόνες από τη θεωρία γράφων, την κοινωνιολογία, την εθνογραφία και τα μαθηματικά. [25]

Η εξόρυξη δεδομένων στα κοινωνικά δίκτυα αντιμετωπίζει αρκετές δυσκολίες σε κάθε εγχείρημά της. Αρκετά συχνά εμφανίζονται τα τρία παράδοξα των μεγάλων δεδομένων (big data paradox) τα οποία είναι τα εξής:

- **Παράδοξο διαφάνειας:** Ενώ η χρήση των μεγάλων δεδομένων κάνει τον κόσμο πιο κατανοητό και διαφανή, η συλλογή αυτών των δεδομένων γίνεται κρυφά και τα εργαλεία που χρησιμοποιούνται είναι αδιαφανή και περιορισμένα από νομικά πλαίσια.
- **Παράδοξο ταυτότητας:** Η πρόσβαση σε μεγάλα δεδομένα που περιέχουν πληροφορίες για τις αγοραστικές συνήθειες, τα τηλεφωνήματα, τις αναρτήσεις στα κοινωνικά δίκτυα μπορούν να επηρεάσουν τις συμπεριφορές των χρηστών προς μια συγκεκριμένη κατεύθυνση. Κάθε χρήστης του Google επηρεάζεται από την ανάδραση των μεγάλων δεδομένων στα αποτελέσματα αναζήτησης που προβάλλονται. Εταιρίες που εξάγουν πληροφορίες από τα δεδομένα θα γνωρίζουν περισσότερα για εμάς σε σχέση με τον ίδιο μας τον εαυτό. Ένα τέτοιο παράδειγμα είναι το γεγονός πως μια προωθητική εταιρία διαπίστωσε πως μια κοπέλα ήταν έγκυος πιο γρήγορα από τον ίδιο της τον πατέρα. [26]
- **Παράδοξο δύναμης:** Στις εξεγέρσεις της αραβικής άνοιξης στην Αίγυπτο το 2010 τα κοινωνικά μέσα δικτύωσης βοήθησαν αρκετά τους διοργανωτές και παίξανε καθοριστικό ρόλο στην οργάνωση όλων των δράσεων. Όμως οι αναλύσεις των μεγάλων δεδομένων από τα κοινωνικά δίκτυα είναι υπό τον έλεγχο μεγάλων ιδρυμάτων και όχι απλών ανθρώπων. Έτσι το καθεστώς στη Συρία και την Τουρκία τοποθέτησε περιορισμούς ή και απαγόρευσε τα κοινωνικά δίκτυα στους πολίτες της. [27]

Εκτός από τα προβλήματα των παραδόξων μπορούν να εμφανιστούν ανεπαρκή, θορυβώδη και εσφαλμένα στοιχεία στη συλλογή των δεδομένων. Το social media mining αναπαριστά τα κοινωνικά μέσα δικτύωσης με έναν υπολογιστικό τρόπο σε μετρήσιμη μορφή ώστε να κατανοήσουμε καλύτερα τις αλληλεπιδράσεις μέσα σε αυτά.

Η εξόρυξη δεδομένων από τα κοινωνικά δίκτυα χρησιμοποιείται σε αρκετούς τομείς και είδη υπηρεσιών όπως: διοίκηση και ανάπτυξη επιχειρήσεων, έρευνα κοινωνικών συμπεριφορών, υπηρεσίες υγείας και εκπαίδευτικά ιδρύματα. Συχνά οι εταιρίες χρησιμοποιούν τα μοτίβα συνδεσιμότητας που διαδέχονται τα κοινωνικά δίκτυα, όπως η κοινωνική ομοιότητα (social similarity) μεταξύ χρηστών που προκαλείται από επιρροή, αμοιβαιότητα και μεταβατικότητα. Οι κοινωνικοί αναλυτές χρησιμοποιούν επίσης την συναισθηματική ανάλυση για να εντοπίσουν θετικά ή αρνητικά συναισθήματα στις αναρτήσεις και τα σχόλια των χρηστών. Αυτό παρέχει πολύ σημαντική πληροφορία για τα συναισθήματα των χρηστών σε συγκεκριμένα θέματα.

Η χρήση των εξαγομένων πληροφοριών όμως δεν περιορίζεται μόνο στην ανάλυση. Με βάση τα εξαγόμενα αποτελέσματα από την ανάλυση μιας ομάδας χρηστών στα κοινωνικά δίκτυα, μπορεί να εντοπισθεί ο χρήστης που ασκεί τη μεγαλύτερη επιρροή στους άλλους. Αυτοί οι χρήστες (Influencers) καθορίζονται από την αναγνωρισμότητα, τη δραστηριότητα, τη γενιά και την πρωτοπορία που έχουν. Πολλοί χρήστες πλέον έχουν αρχίσει να δίνουν ολοένα και μεγαλύτερη βαρύτητα στις γνώμες άλλων χρηστών για να λύσουν ένα πρόβλημα ή να κατανοήσουν ένα αντικείμενο. Έτσι οι εταιρίες μπορούν και συστήνουν συγκεκριμένα προϊόντα σε συγκεκριμένο αγοραστικό κοινό και αντίστοιχα τα μέσα κοινωνικής δικτύωσης μπορούν και προτείνουν νέες φιλίες με άλλους χρήστες και εμφανίζουν προτεινόμενες σελίδες.

3.2.2. Facebook Graph API

Το Facebook αναμφισβήτητα είναι η καρδιά των κοινωνικών μέσων δικτύωσης και το αποδεικνύει με τους παραπάνω από 2.2 δισεκατομμύρια χρήστες. Το Facebook θεωρείται ως ένας κοινωνικός γράφος, ο οποίος σταδιακά μετατρέπεται σε γράφο ενδιαφέροντος καθώς ο γράφος αυτός περιέχει πληροφορίες μέσω των σελίδων και των Likes για τις σχέσεις μεταξύ των χρηστών και τα πράγματα που τους “ενδιαφέρουν”. Έτσι εύκολα θα μπορούσε να χαρακτηριστεί ως γράφος κοινωνικού ενδιαφέροντος (social interest graph).

Από τη σκοπιά του social media mining τα δεδομένα που είναι αποθηκευμένα στο Facebook για άτομα, ομάδες και προϊόντα είναι αρκετά συναρπαστικά και με μεγάλη δυναμική, λόγω του πολύ καλά δομημένου API που προσφέρει το Facebook για να εξάγει κανείς τα δεδομένα και να τα συνθέσει σε πληροφορία.

API (Application Programming Interface) ορίζουμε μια διεπαφή μεταξύ προγραμματιστικών διεπαφών που παρέχει ένα λειτουργικό σύστημα, βιβλιοθήκη ή εφαρμογή προκειμένου να επιτρέπει να γίνονται προς αυτό αιτήσεις (queries) από άλλα προγράμματα και ανταλλαγή δεδομένων. Ένας από τους βασικούς σκοπούς μίας διεπαφής είναι να ορίζει και να διατυπώνει το σύνολο των λειτουργιών-υπηρεσιών που μπορεί να παρέχει μια βιβλιοθήκη ή ένα λειτουργικό σύστημα σε άλλα προγράμματα, χωρίς να επιτρέπει πρόσβαση στον κώδικα που υλοποιεί αυτές τις υπηρεσίες [28]

Το Graph API είναι ο κύριος τρόπος να συγκεντρώσει κανείς δεδομένα από την πλατφόρμα του Facebook. Είναι μια χαμηλού επιπέδου HTTP διεπαφή την οποία χρησιμοποιούν οι εφαρμογές για να κάνουν αιτήσεις δεδομένων, να ανεβάσουν περιεχόμενο, να διαχειριστούν διαφημίσεις και πολλές άλλες εργασίες. Έτσι κατά κύριο λόγο οι εφαρμογές διαβάζουν αλλά και γράφουν τον κοινωνικό γράφο.

Η πλατφόρμα του Facebook είναι μια πολύ καλά δομημένη και εύχρηστη δίοδος στη μεγαλύτερη αποθήκη δεδομένων σε έκταση και βάθος πληροφορίας. Το μοντέλο φιλίας στο

Facebook είναι συμμετρικό, που σημαίνει πως απαιτεί κοινή συναίνεση μεταξύ χρηστών για να μπορεί να έχει ο ένας πρόσβαση στις ιδιωτικές διαδράσεις του άλλου. Ένας αναλυτής για να έχει πρόσβαση στα δεδομένα ενός χρήστη πρέπει πρώτα να δηλώσει μια εφαρμογή που θα χρησιμοποιεί το API και να χρησιμοποιήσει την εφαρμογή αυτή ως αφετηρία για την πλατφόρμα των προγραμματιστών του Facebook. Τα μόνα δεδομένα που θα είναι διαθέσιμα στην εφαρμογή είναι αυτά που οι χρήστες έχουν δώσει την άδειά τους.

Το “σύμβολο πρόσβασης” (access token) που εμφανίζεται στην εφαρμογή είναι ένα OAuth (Open Authorization) token που παρέχεται ως ασφάλεια για τους συνδεδεμένους χρήστες. Το πρωτόκολλο του OAuth σημαίνει πως επιτρέπεται στους χρήστες να δώσουν άδεια σε εφαρμογές τρίτων, οι οποίες θα έχουν πρόσβαση στα δεδομένα χωρίς να χρειάζεται να μοιρασθούν ευαίσθητες πληροφορίες π.χ. κωδικός πρόσβασης.

The screenshot shows the Graph API Explorer interface. At the top, there is a header with 'Graph API Explorer', 'Home > Tools > Graph API Explorer', and an 'Application:' dropdown. Below the header, there is a text input for 'Access Token' containing a long string of characters, and a 'Debug' button. Underneath, there are tabs for 'Graph API' (which is selected) and 'FQL Query'. A 'GET' dropdown menu is open, showing the URL '/644382747?fields=id,name,friends.fields(likes)'. To the right of the URL, there is a 'Learn more about new features.' link. The main area is divided into two sections: 'Node: 644382747' on the left and the JSON response on the right. The 'Node' section shows checkboxes for selecting fields: 'id', 'name', 'friends' (with a dropdown arrow), and 'likes'. The 'friends' and 'likes' checkboxes are checked. The JSON response on the right shows the following data:

```

{
  "id": "644382747",
  "name": "Matthew A. Russell",
  "friends": {
    "data": [
      {
        "id": "6224364"
      },
      {
        "id": "51801220",
        "likes": {
          "data": [
            {
              "category": "Sports/recreation/activities",
              "category_list": [
                {
                  "id": "184405378265823",
                  "name": "Gym"
                }
              ],
              "name": "SOF WODs",
              "id": "226640360698227",
              "created_time": "2013-04-03T17:50:41+0000"
            },
            {
              "category": "Local business",
              "category_list": [
                {
                  ...
                }
              ]
            }
          ]
        }
      }
    ]
  }
}

```

Εικόνα 3.3. Η εφαρμογή του Graph API με ένα αίτημα για τα ενδιαφέροντα των φίλων.

3.3 Ανάλυση των Δεδομένων

Ανάλυση δεδομένων είναι η διαδικασία επιθεώρησης, καθαρισμού, μοντελοποίησης των δεδομένων με στόχο την ανακάλυψη χρήσιμων πληροφοριών και συμπερασμάτων. Ο τομέας της ανάλυσης έχει πολλές πλευρές και προσεγγίσεις καθώς συμπεριλαμβάνει ποικίλες τεχνικές. Σε στατιστικές εφαρμογές, η ανάλυση διαχωρίζεται σε περιγραφική στατιστική (descriptive analysis),

ερευνητική ανάλυση(exploratory data analysis EDA) και επαληθευτική ανάλυση (confirmatory data analysis CDA). [29]

Η ερευνητική ανάλυση είναι το πρώτο στάδιο της διαδικασίας ανάλυσης. Σε αυτό το στάδιο διευκρινίζεται ο χειρισμός των δεδομένων, θέτονται τα κατάλληλα ερωτήματα και καθορίζεται ο καλύτερος τρόπος παρουσίασης των συμπερασμάτων. Εδώ ο αναλυτής εκμαileύει τάσεις, μοτίβα και μη αναμενόμενα αποτελέσματα από το μοντέλο χρησιμοποιώντας ποσοτικές και παραστατικές μεθόδους. Η θεμέλια δομή των δεδομένων, η αναγνώριση λαθών-ανωμαλιών, τα ελλιπή στοιχεία και ο έλεγχος αρχικών υποθέσεων γίνονται επίσης σε αυτό το στάδιο.

Στην επαληθευτική ανάλυση αξιολογούνται τα στοιχεία χρησιμοποιώντας παραδοσιακά στατιστικά εργαλεία. Ουσιαστικά εδώ ο αναλυτής δοκιμάζει τις αρχικές του υποθέσεις. Ένα μεγάλο μέρος της επαληθευτικής ανάλυσης είναι η ποσοτικοποίηση των δεδομένων, ώστε να ελεγχθεί η εγκυρότητα του μοντέλου ανάλυσης. [30]

Η ανάλυση αναφέρεται στον κατακερματισμό του συνόλου σε διαφορετικά συστατικά για μεμονωμένη εξέταση. Μέσα από αυτή τη διαδικασία αντλούνται τα αρχικά ανεπεξέργαστα δεδομένα και μετατρέπονται σε χρήσιμη πληροφορία για σωστή λήψη αποφάσεων. Απαντώνται ερωτήματα, δοκιμάζονται υποθέσεις και διαψεύδονται θεωρίες.

Η συνολική διαδικασία της ανάλυσης δεδομένων μπορεί να αναλυθεί σε 7 διακριτές φάσεις:

1. **Προϋποθέσεις Δεδομένων:** Τα δεδομένα συγκεκριμένοποιούνται με βάση τις προϋποθέσεις που θέτουν οι αναλυτές ή οι πελάτες. Συγκεκριμένες μεταβλητές που σχετίζονται π.χ. με τον πληθυσμό (ηλικία, εισόδημα) πρέπει πάντα να είναι στην ίδια συγκεκριμένη μορφή π.χ. (αριθμητική ή κατηγορηματική).
2. **Συγκέντρωση Δεδομένων:** Τα δεδομένα συγκεντρώνονται από διάφορες πηγές, όπως: αισθητήρες στο περιβάλλον, κάμερες ασφαλείας, δορυφόρους κτλ. πέρα από διαδικτυακές πηγές.
3. **Επεξεργασία Δεδομένων:** Τα αρχικά δεδομένα πρέπει να επεξεργαστούν και να οργανωθούν για να αναλυθούν. Για παράδειγμα, τα δεδομένα τοποθετούνται σε στήλες και γραμμές σε υπολογιστικό φύλλο.
4. **Εκκαθάριση Δεδομένων:** Αφού οργανώσει ο αναλυτής τα δεδομένα, πρέπει να ελέγχει αν τα δεδομένα είναι ελλιπή, αν περιέχονται διπλότυπα ή και αν είναι εσφαλμένα. Ασυνήθιστες τιμές για παράδειγμα, οι οποίες είναι πιο πάνω ή κάτω από τα προκαθορισμένα όρια θα πρέπει να αξιολογηθούν μεμονωμένα. Υπάρχουν επίσης

εργαλεία όπως ο έλεγχος ορθογραφίας για δεδομένα κειμένου, τα οποία μειώνουν τα λάθη στις λέξεις.

5. **Ερευνητική Ανάλυση Δεδομένων:** Εδώ εφαρμόζονται τεχνικές ανάλυσης στα δεδομένα ώστε ο αναλυτής να αρχίσει να καταλαβαίνει το μήνυμα που περιέχεται στα δεδομένα. Περιγραφική στατιστική, όπως ο μέσος όρος ή ο διάμεσος βοηθούν στην καλύτερη κατανόηση.
6. **Μοντελοποίηση και Αλγόριθμοι:** Μαθηματικές φόρμουλες και μοντέλα εφαρμόζονται στα δεδομένα ώστε να αναγνωριστούν σχέσεις μεταξύ των μεταβλητών των δεδομένων όπως συσχετίσεις και αιτιότητες. Με μαθηματικούς όρους η συμπερασματική στατιστική θέτει μια εξαρτημένη Y και μια ανεξάρτητη X μεταβλητή όπου, το Y είναι συνάρτηση του X . Για παράδειγμα, οι πωλήσεις(Y) είναι συνάρτηση της διαφήμισης(X). Καταλήγουμε σε ένα μοντέλο $Y = aX + b + \text{error}$.
7. **Επικοινωνία:** Αφού τα δεδομένα έχουν αναλυθεί, πρέπει να γίνει η αναφορά τους στο κοινό. Ένας κλασσικός τρόπος για να επικοινωνήσει κανείς τα αποτελέσματα της ανάλυσης είναι μέσω τεχνικών οπτικοποίησης (Data Visualization). Έτσι το τελικό μήνυμα μεταφέρεται ευκολότερα και γίνεται σαφέστερο προς το κοινό.

3.4 Μηχανική Μάθηση

Η μηχανική μάθηση είναι μία μέθοδος της ανάλυσης δεδομένων που αυτοματοποιεί το χτίσιμο μοντέλων ανάλυσης. Είναι κομμάτι της τεχνητής νοημοσύνης βασισμένο στην ιδέα ότι τα συστήματα μαθαίνουν από τα δεδομένα, αναγνωρίζουν επαναλαμβανόμενα μοτίβα και παίρνουν αποφάσεις με την ελάχιστη ανθρώπινη παρέμβαση. Τα αναλυτικά μοντέλα επιτρέπουν στους ερευνητές, στους επιστήμονες δεδομένων και στους αναλυτές να παράγουν αξιόπιστες αποφάσεις και αποτελέσματα και να αναδείξουν αλληλοσυσχετίσεις μέσω μάθησης από ιστορικές σχέσεις και τάσεις στα δεδομένα.

Λόγω των νέων τεχνολογιών, η μηχανική μάθηση του σήμερα διαφέρει αρκετά από τη μηχανική μάθηση των προηγούμενων ετών. Προέρχεται από την αναγνώριση μοτίβων και τη θεωρία ότι οι υπολογιστές μπορούν να εκπαιδευτούν χωρίς να είναι προγραμματισμένοι να κάνουν συγκεκριμένα καθήκοντα.

Αν και οι αλγόριθμοι μηχανικής μάθησης υπάρχουν από καιρό, η ικανότητα να εφαρμοστούν αυτόματα σύνθετες μαθηματικές φόρμουλες σε μεγάλα δεδομένα και να παραχθούν επανειλημμένα γρήγορα αποτελέσματα είναι μια πρόσφατη εξέλιξη. Χτίζοντας μοντέλα με μεγάλη ακρίβεια, ένας

οργανισμός έχει πολύ μεγαλύτερες πιθανότητες να αναγνωρίσει ευκαιρίες για κέρδος ή να αποφύγει κινδύνους και ρίσκα. [31]

Μερικές πολύ γνωστές εφαρμογές της μηχανικής μάθησης είναι :

- Το αυτοκίνητο της Google που οδηγείται μόνο του.
- Τα συστήματα που προτείνουν και συνιστούν προϊόντα ανάλογα με τις προηγούμενες αγορές
- Ανάλυση γνώμης πελατών στα κοινωνικά δίκτυα
- Αναγνώριση απάτης

3.4.1 Τι ορίζουμε ως μηχανική μάθηση

Πρώτος ο Arthur Samuel το 1959 όρισε τη μηχανική μάθηση ως το πεδίο επιστήμης που δίνει στους υπολογιστές τη δυνατότητα να μαθαίνουν χωρίς να είναι αποκλειστικά προγραμματισμένοι. Αρκετά χρόνια αργότερα, το 1997 ο Tom Mitchell έδωσε έναν “καλά τοποθετημένο” ορισμό για την μηχανική μάθηση που χρησιμοποιείται μέχρι σήμερα.

Σύμφωνα με τον Mitchell, ένα υπολογιστικό πρόγραμμα στο οποίο έχει ανατεθεί η μάθηση από μια εμπειρία E με αναφορά σε κάποιο καθήκον T και κάποια μέτρηση επίδοσης P , αν η επίδοση στο T , όπως μετρήθηκε από το P , βελτιώνεται με την εμπειρία E . Για παράδειγμα, αν θέλουμε ένα πρόγραμμα να προβλέπει τα μοτίβα στην κυκλοφορία οχημάτων σε ένα σημείο τομής (καθήκον T) π.χ. διασταύρωση, τότε τρέχουμε έναν αλγόριθμο μηχανικής μάθησης με δεδομένα από προηγούμενα μοτίβα και μετρήσεις (εμπειρία E) και αν έχει μάθει επιτυχώς, τότε θα υπάρξει βελτίωση στις προβλέψεις του μέλλοντος (μέτρηση απόδοσης P).

Οι μέθοδοι μηχανικής μάθησης ταξινομούνται σε 3 μεγάλες κατηγορίες :

- **Επιβλεπόμενη μάθηση (Supervised Learning):** Το υπολογιστικό πρόγραμμα δέχεται παραδείγματα ως εισόδους καθώς και τα επιθυμητά αποτελέσματα. Στόχος είναι να μάθει έναν γενικό κανόνα προκειμένου να αντιστοιχίσει και να συγκρίνει τις εισόδους με τα αποτελέσματα και να παράξει ακριβή συμπεράσματα σε νέες εισόδους. Μέσα από μεθόδους όπως η ομαδοποίηση (clustering), η οπισθοδρόμηση (regression), πρόβλεψη (prediction), η επιβλεπόμενη μάθηση χρησιμοποιεί τα μοτίβα για να προβλέψει τις τιμές στις ετικέτες σε δεδομένα τα οποία δεν έχουν κάποια ετικέτα.
- **Μη Επιβλεπόμενη μάθηση (Unsupervised Learning):** Χωρίς να παρέχεται κάποια εμπειρία στον αλγόριθμο μάθησης, πρέπει να βρει τη δομή των δεδομένων εισόδου. Το σύστημα δεν ξέρει τη σωστή απάντηση και πρέπει να βρει τί του έχει διθεί ως είσοδος. Ανακαλύπτει δηλαδή κρυμμένα

μοτίβα σε δεδομένα. Η μη επιτηρούμενη μάθηση δουλεύει καλά σε δεδομένα συναλλαγών γιατί μπορεί να αναγνωρίσει στοιχεία πελατών με παρόμοιες συμπεριφορές και μετρήσεις, οι οποίοι μετέπειτα μπορούν να αντιμετωπίζονται όμοια σε εκστρατείες μάρκετινγκ.

- **Ενισχυτική μάθηση (Reinforcement Learning):** Ο αλγόριθμος ανακαλύπτει μέσω δοκιμών και λαθών ποιες ενέργειες προκαλούν τα καλύτερα αποτελέσματα. Αυτή η μέθοδος έχει 3 βασικά στοιχεία: τον πράκτορα (agent), το περιβάλλον (όλα με όσα διαδρά ο πράκτορας) και οι δράσεις ενέργειες (αυτά που μπορεί να κάνει ο πράκτορας). Η αποστολή για τον πράκτορα είναι είναι να επιλέγει ενέργειες, οι οποίες μεγιστοποιούν την ανταμοιβή σε ένα συγκεκριμένο χρονικό διάστημα. Ο πράκτορας θα φτάσει πολύ γρηγορότερα αν ακολουθήσει μια καλή πολιτική. Έτσι στόχος είναι για τον πράκτορα να βρει την καλύτερη πολιτική. [32] [33]

3.4.2. Επιβλεπόμενη μάθηση (Supervised Learning)

Στην πλειονότητα των εφαρμογών επιβλεπόμενης μάθησης, ο σκοπός είναι να αναπτυχθεί μια συνάρτηση $h(x)$ που λέγεται υπόθεση. Η μάθηση συμπεριλαμβάνει εκλεπτυσμένους μαθηματικούς αλγόριθμους για να βελτιστοποιήσει τη συνάρτηση, έτσι ώστε με αρχικά δεδομένα x για ένα συγκεκριμένο τομέα να παραχθεί μια ακριβής εκτίμηση $h(x)$.

Για παράδειγμα, ένα σύστημα που προβλέπει την τιμή ενός σπιτιού δεν δέχεται ως είσοδο μόνο τα τετραγωνικά μέτρα (x_1) αλλά και τον αριθμό δωματίων (x_2), τον αριθμό μπάνιων (x_3), τον αριθμό ορόφων (x_4), το έτος κατασκευής (x_5) και πολλά άλλα. Ο καθορισμός των κατάλληλων εισόδων είναι σημαντικό κομμάτι της Μηχανικής μάθησης.

Έστω ένα απλό σύστημα πρόβλεψης έχει τη μορφή:

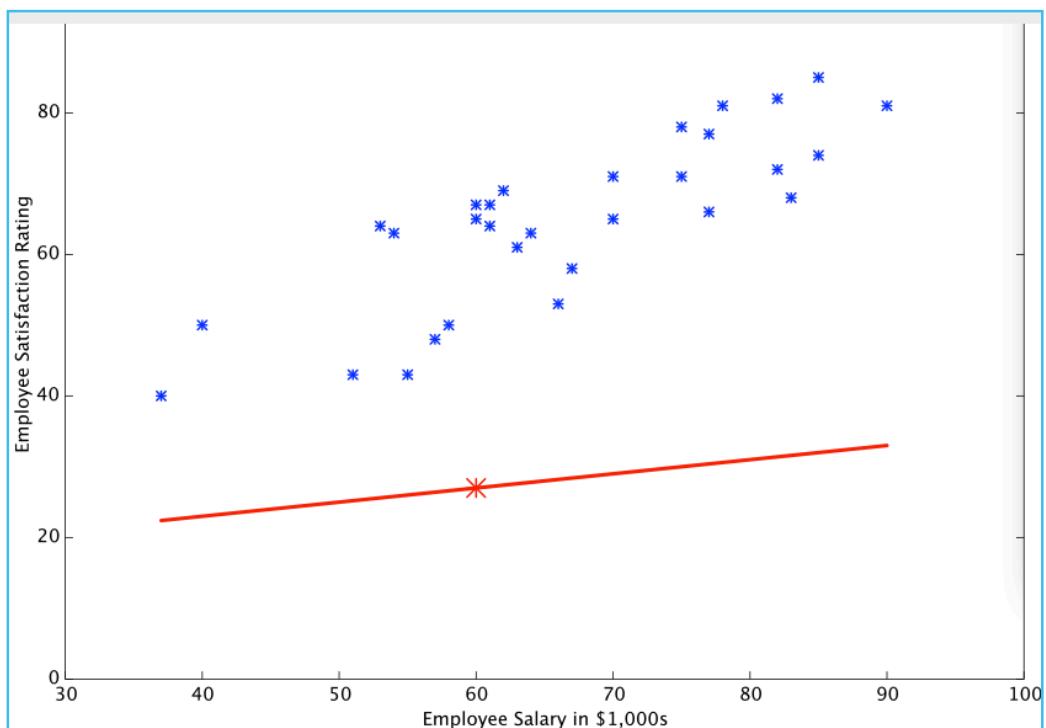
$$h(x) = \theta_0 + \theta_1 x$$

όπου θ_0 και θ_1 είναι σταθερές.

Στόχος είναι να βρεθούν οι καταλληλότερες τιμές των θ_0 και θ_1 για να δουλέψει το σύστημα όσο καλύτερα γίνεται. Η βελτιστοποίηση της συνάρτησης $h(x)$ γίνεται μέσα από παραδείγματα εκπαίδευσης (training examples). Για κάθε παράδειγμα εκπαίδευσης έχουμε μία είσοδο x_{train} , για την οποία η αντίστοιχη έξοδος γίνεται γνωστή. Για κάθε παράδειγμα, βρίσκουμε τη διαφορά μεταξύ της γνωστής τιμής για της προβλεπόμενης τιμής $h(x_{train})$. Με αρκετά παραδείγματα εκπαίδευσης, αυτές οι διαφορές μας δίνουν έναν χρήσιμο τρόπο μέτρησης “σφαλματότητας” (wrongness) της $h(x)$. Μεταβάλλοντας τις τιμές των θ_0 και θ_1 μεταβάλλουμε και την τιμή της $h(x)$ λιγότερο εσφαλμένα. Αυτή η διαδικασία επαναλαμβάνεται ξανά και ξανά μέχρι το σύστημα να συγκλίνει στις καλύτερες τιμές για θ_0 και θ_1 . Κατά αυτό τον τρόπο το σύστημα εκπαιδεύεται και είναι έτοιμο να κάνει προβλέψεις πραγματικού κόσμου.

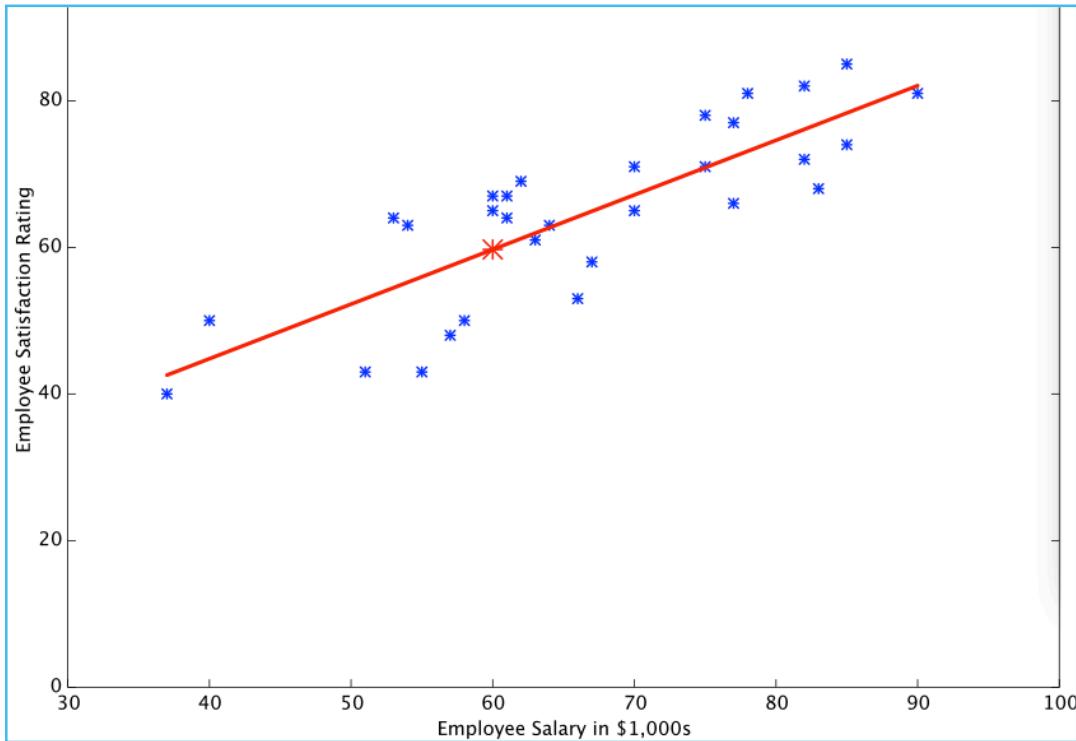
Ο βρετανός μαθηματικός και στατιστικολόγος George E. Box είπε πως “όλα τα μοντέλα είναι εσφαλμένα, κάποια όμως είναι χρήσιμα”. Η μηχανική μάθηση βασίζεται αρκετά στη στατιστική. Για παράδειγμα, όταν εκπαιδεύουμε το σύστημα να μαθαίνει, πρέπει να έχουμε αρκετά τυχαίες εισόδους ως δεδομένα εκπαίδευσης. Αν τα δεδομένα εκπαίδευσης δεν είναι τυχαία, υπάρχει ο κίνδυνος να βρεθούν μοτίβα μηχανικής μάθησης που στην πραγματικότητα δεν υπάρχουν. Επίσης αν τα δεδομένα εκπαίδευσης είναι λίγα, το σύστημα δεν μπορεί να εκπαιδευτεί καλά με αποτέλεσμα να παράγει λάθος συμπεράσματα.

Για παράδειγμα, σε ένα σύνολο δεδομένων εκπαίδευσης με μισθούς εργαζομένων και τον αντίστοιχο βαθμό ικανοποίησης, βλέπουμε πως τα δεδομένα είναι αρκετά θορυβώδη “noisy”, όμως ακολουθούν μια λογική πορεία. Όσο αυξάνεται ο μισθός, αυξάνεται και η ικανοποίηση του εργαζομένου. Με αυτό το σκεπτικό αρχικοποιούμε τη συνάρτηση $h(x)$ με κάποιες τιμές για τις σταθερές θ_0 και θ_1 . Στις πρώτες προβλέψεις του συστήματος βλέπουμε πως το σύστημα δε γνωρίζει αρκετά και δεν είναι ακόμα σωστά εκπαιδευμένο.



Εικόνα 3.5. Η πρώτη πρόβλεψη του συστήματος για $h(x) = 12 + 0.2x$

Έπειτα από αρκετές επαναλήψεις και ελέγχους για τις τιμές των σταθερών, το εκπαιδευμένο πλέον σύστημα θα συγκλίνει σε κάποιες συγκεκριμένες τιμές για τις σταθερές. Τότε η συνάρτηση $h(x)$ είναι η βέλτιστη.



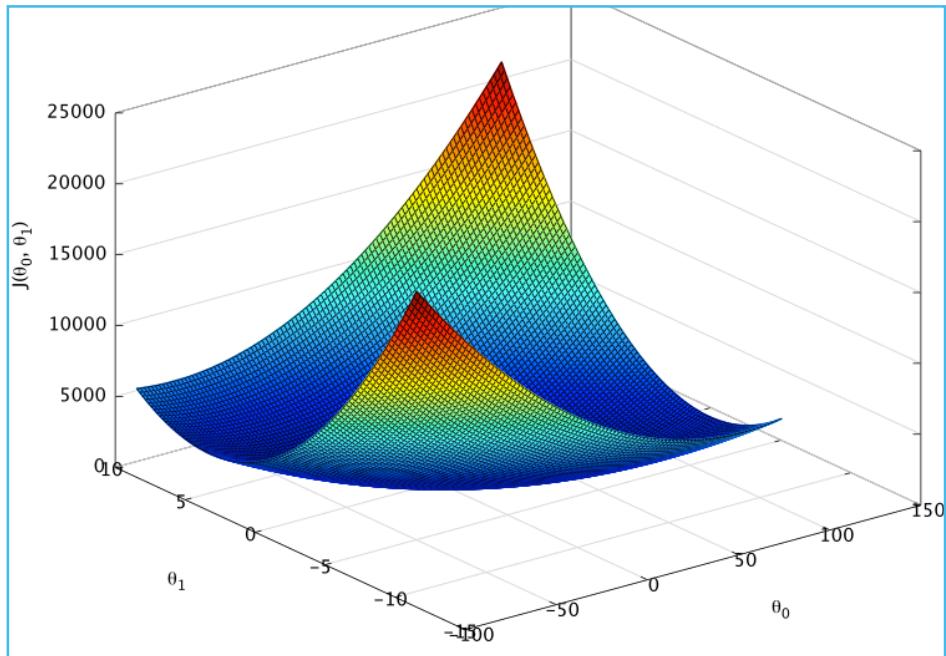
Εικόνα 3.6. Η βέλτιστη συνάρτηση που παράγει το σύστημα.

Η μέτρηση σφάλματος (wrongness measure) είναι γνωστή ως συνάρτηση κόστους ή συνάρτηση απώλειας και συμβολίζεται με $J(\theta)$. Το όρισμα θ αναπαριστά όλους τους συντελεστές στη συνάρτηση πρόβλεψης. Τα περισσότερα προβλήματα μηχανικής μάθησης στον πραγματικό κόσμο δέχονται εκατοντάδες έως εκατομμύρια διαστάσεις δεδομένων για να μπορέσουν να κάνουν προβλέψεις χρησιμοποιώντας χιλιάδες σταθερές. Για παράδειγμα ένα σύστημα $h(x)$ που δέχεται 4 μεταβλητές τις x_1, x_2, x_3, x_4 θα είναι της μορφής:

$$h(x_1, x_2, x_3, x_4) = \theta_0 + \theta_1 x_1 + \theta_2 x_2^2 + \theta_3 x_3 x_4 + \theta_4 x_1^3 x_2^2 + \theta_5 x_2 x_3^4 x_4^2$$

Έτσι σε μια απλούστερη περίπτωση το θ είναι συνδυασμός όλων των σταθερών και το $J(\theta_0, \theta_1, \theta_2, \theta_3, \theta_4)$ μας δίνει μια μαθηματική μέτρηση πόσο λάθος ήταν το σύστημα μας με τις συγκεκριμένες εισόδους.

Με τη μέθοδο ελάχιστων τετραγώνων, η ποινή για λάθος προβλέψεις αυξάνεται τετραγωνικά με τη διαφορά ανάμεσα στην πρόβλεψη και στην σωστή απάντηση. Στόχος είναι να βρεθεί η ελάχιστη τιμή της συνάρτησης J .



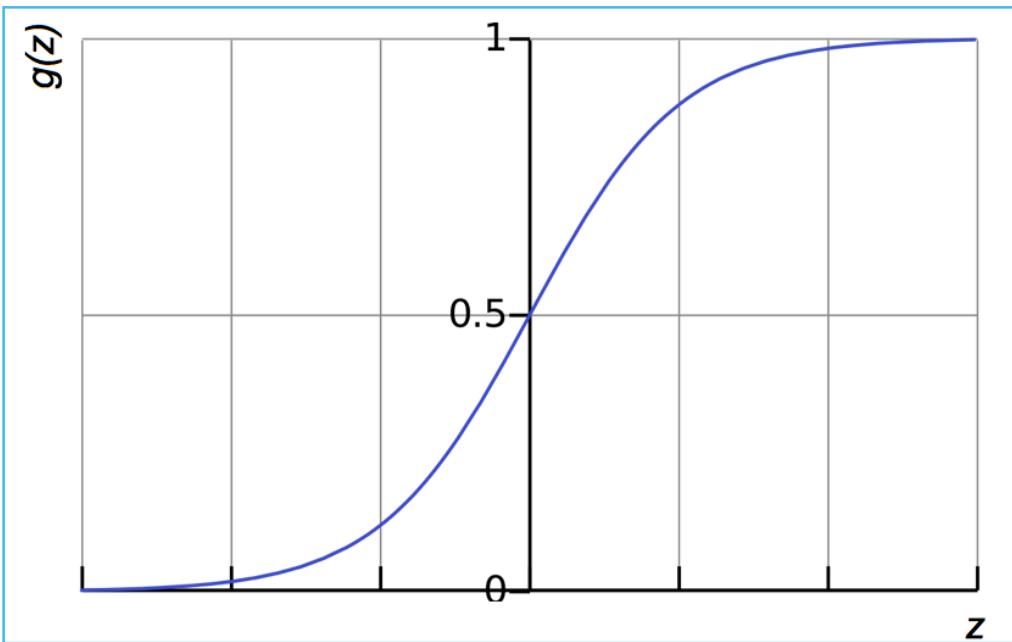
Εικόνα 3.7. Γραφική αναπαράσταση της συνάρτησης κόστους.

Βλέπουμε πως η γραφική παράσταση σχηματίζει κάτι σαν λεκάνη. Ο πυθμένας του γραφήματος αναπαριστά την ελάχιστη τιμή της J . Κάθε φορά ο αλγόριθμος εξετάζει τιμές για τις σταθερές, αφαιρώντας λίγο από τη μία και προσθέτοντας λίγο στην άλλη. Αν το αποτέλεσμα είναι καλύτερο και έχουμε κατέβει χαμηλότερα στη λεκάνη, τότε έχει ολοκληρωθεί επιτυχώς ένας γύρος μάθησης.

Το παραπάνω παράδειγμα ανήκει στην κατηγορία της παρεμβολής (Regression), όπου χρησιμοποιούμε τη μηχανική μάθηση για να απαντήσουμε ερωτήσεις με συνεχή ποσοτικά στοιχεία. Δηλαδή πού βρίσκεται η τιμή της πρόβλεψης σε ένα συνεχόμενο φάσμα. Π.χ. πρόβλεψη θερμοκρασίας για κάθε μέρα της επόμενης βδομάδας.

Παράλληλα με την παρεμβολή υπάρχει και μια δεύτερη μεγάλη υποκατηγορία εφαρμογών, η ταξινόμηση (Classification). Στην ταξινόμηση θέλουμε η συνάρτηση πρόβλεψης να κάνει προβλέψεις από ένα πεπερασμένο διακριτό σύνολο τιμών.

Αυτήν τη συμπεριφορά αναπαριστά η συνάρτηση Sigmoid $g(z)$ και ισχύει $h(x) = g(z)$. Το z είναι μια αναπαράσταση της εισόδου και των σταθερών, με $z = \theta_0 + \theta_1 x$ και αντίστοιχα $h(x) = g(\theta_0 + \theta_1 x)$.



Εικόνα 3.8. Γραφική αναπαράσταση της συνάρτησης $g(z)$.

Η λογική στη συνάρτηση κόστους διαφέρει στην ταξινόμηση σε σχέση με την παρεμβολή. Αν η σωστή απάντηση γίνεται 0 και το σύστημα προέβλεψε 1, τότε είναι τελείως λάθος και το κόστος θα πρέπει να είναι αρκετά μεγάλο. Το ίδιο και στην ανάποδη περίπτωση. Αν όμως η πρόβλεψη γίνεται σωστή αλλά όχι τελείως σίγουρη ($y = 1$ και $h(x) = 0.8$) τότε θα πρέπει να επιστραφεί ένα μικρό κόστος. Αντιθέτως αν η πρόβλεψη γίνεται λάθος αλλά όχι τελείως σίγουρη ($y = 1$ και $h(x) = 0.3$) τότε θα επιστραφεί μεγάλο κόστος αλλά όχι τόσο μεγάλο αν γίνεται λάθος. [34] [36]

Αυτήν τη συμπεριφορά την αναπαριστά η λογαριθμική συνάρτηση :

$$cost = \begin{cases} -\log(h(x)), & \text{if } y = 1 \\ -\log(1 - h(x)), & \text{if } y = 0 \end{cases}$$

3.4.3 Μη επιβλεπόμενη Μάθηση (Unsupervised Learning)

Η μη επιβλεπόμενη μάθηση βοηθά στην ανακάλυψη μέχρι πρότινος άγνωστων μοτίβων σε ένα σετ δεδομένων, όμως τα περισσότερα από αυτά τα μοτίβα που προβλέπονται είναι μερικώς εσφαλμένα. Μη γνωρίζοντας ποιο θα είναι το αποτέλεσμα, δεν υπάρχει τρόπος να αναγνωριστεί πόσο ακριβές θα είναι οι προβλέψεις. Το καλύτερο πεδίο εφαρμογής είναι όταν τα δεδομένα δεν έχουν κάποια επιθυμητά αποτελέσματα, όπως για παράδειγμα ο καθορισμός του αγοραστικού κοινού για ένα προϊόν που βγαίνει για πρώτη φορά στην αγορά.

Μερικές από τις εφαρμογές της μη επιβλεπόμενης μάθησης είναι:

- **Ομαδοποίηση (Clustering):** Χωρίζει το σετ δεδομένων σε υποομάδες ανάλογα με τις ομοιότητες που εμφανίζουν.
- **Αναγνώριση Ανωμαλιών (Anomaly detection):** Αυτόματα ανακαλύπτονται ασυνήθιστες τιμές μέσα στα δεδομένα. Έτσι αναγνωρίζονται ευκολότερα οι υποκλοπές, τα ελαττωματικά κομμάτια στο hardware και τα σφάλματα που προκαλούνται από ανθρώπινες παρεμβάσεις.
- **Εξόρυξη συσχετίσεων (Association Mining):** Αναγνωρίζει ομάδες δεδομένων που εμφανίζονται συχνά στα δεδομένα εισόδου. Χρησιμοποιείται ευρέως στην ανάλυση καλαθιού (basket analysis), ώστε να μπορούν οι καταστηματάρχες εύκολα να ανακαλύψουν τις συνδυαστικές αγορές των πελατών.
- **Μοντέλο λανθάνων μεταβλητών (Latent variable models):** Χρησιμοποιείται για την προεπεξεργασία των δεδομένων όπως η μείωση διαστάσεων στο dataset και η αποσύνθεση του dataset σε πολλά κομμάτια. [35] [36]

3.4.4 Ενισχυτική Μάθηση (Reinforcement Learning)

Στην ενισχυτική μάθηση ο πράκτορας πρέπει μόνος του να αποφασίσει πως θα εκτελέσει μια ενέργεια. Λόγω της έλλειψης δεδομένων-παραδειγμάτων εκπαίδευσης, ο πράκτορας εκπαιδεύεται μέσα από δοκιμές και σφάλματα με στόχο τη μεγιστοποίηση της μακροχρόνιας αμοιβής.

Ο πράκτορας αλληλοεπιδρά με το περιβάλλον του σε βήματα διακριτού χρόνου. Σε κάθε βήμα t , ο πράκτορας λαμβάνει μια παρατήρηση o_t , η οποία περιλαμβάνει μια ανταμοιβή r_t . Τότε επιλέγει μια δράση a_t από ένα σύνολο διαθέσιμων δράσεων. Το περιβάλλον αλλάζει σε μια νέα κατάσταση s_{t+1} και η ανταμοιβή r_{t+1} που προέρχεται από την μετάβαση (s_t, a_t, s_{t+1}) είναι προκαθορισμένη. Στόχος της ενισχυτικής μάθησης είναι να συγκεντρώσει όσο το δυνατόν μεγαλύτερη ανταμοιβή.

Η ενισχυτική μάθηση απαιτεί έναν έξυπνο μηχανισμό εξερεύνησης. Οι τυχαίες ενέργειες, χωρίς αναφορά σε κάποια υπολογισμένη κατανομή πιθανοτήτων, έχει αρκετά μικρή απόδοση. Μια αρκετά χρησιμοποιούμενη μέθοδος είναι η e-greedy, όπου ο πράκτορας επιλέγει την ενέργεια την οποία θεωρεί πως θα του επιφέρει το καλύτερο μακροχρόνιο αποτέλεσμα με πιθανότητα $1 - e$. Αν δεν μπορεί να βρεθεί μια τέτοια ενέργεια, τότε ο πράκτορας επιλέγει τυχαία. [36]

Σύνοψη Κεφαλαίου

Σήμερα περισσότερο από ποτέ τα δεδομένα παίζουν ένα πολύ σημαντικό ρόλο και ισχύουν ως προϋπόθεση για την επίλυση πολλών προβλημάτων. Οι αναλυτές αντλούν δεδομένα από πηγές και στη συνέχεια να τα “καθαρίσει”, να τα ταξινομήσει και να τα αναλύσει ώστε να παραχθεί

πληροφορία – γνώση. Επειδή πλέον τα δεδομένα έχουν αρκετά μεγάλο όγκο ονομάζονται πια Big Data και περιγράφονται από τον όγκο, την ποικιλία, την ταχύτητα και την εγκυρότητα τους.

Το πρώτο βήμα για να παραχθεί γνώση και να λυθούν προβλήματα είναι η εξόρυξη των αρχικών δεδομένων. Έτσι συλλέγουμε δεδομένα που συνήθως βρίσκονται στον παγκόσμιο ιστό. Έπειτα γίνεται η επεξεργασία – καθαρισμός των δεδομένων. Τέλος εφαρμόζονται ορισμένες μέθοδοι ώστε να αναλυθούν σωστά τα δεδομένα. Η ταξινόμηση και η ομαδοποίηση είναι οι συνηθέστερες από αυτές τις μεθόδους. Τα αποτελέσματα της ανάλυσης μπορούν να οπτικοποιηθούν για να γίνουν σαφέστερες και πιο κατανοητές οι λεπτομέρειες προς τους αναλυτές αλλά και στο ευρύ κοινό.

Χρησιμοποιώντας αλγορίθμους μηχανικής μάθησης, η ανάλυση των δεδομένων γίνεται ολοένα και πιο ακριβής, γρήγορη, και αποφεύγει αρκετά λάθη και παρερμηνείες. Είτε δώσουμε στον αλγόριθμο ένα αρχικό σύνολο παραδειγμάτων εκπαίδευσης (Επιβλεπόμενη μάθηση), είτε αφήσουμε τον αλγόριθμο να εκπαιδευτεί μόνος του (Μη επιβλεπόμενη μάθηση), αναγνωρίζοντας ο ίδιος τα επαναλαμβανόμενα πρότυπα και σχέσεις μεταξύ των δεδομένων, το αποτέλεσμα – στόχος είναι οι συνεχώς ακριβέστερες προβλέψεις.

Κεφάλαιο 4

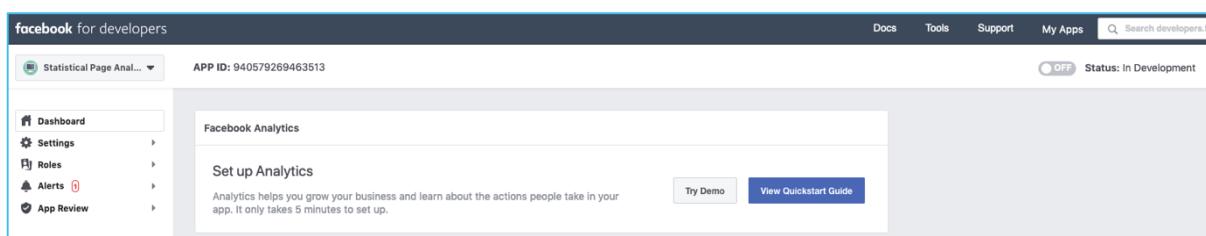
Ανάλυση Δεδομένων Κοινωνικών Δικτύων

4.1 Δεδομένα από Facebook Page

Για την δημιουργία της εφαρμογής και την εκπόνηση της εργασίας ήταν αναγκαίο πρωτίστως η δημιουργία μίας σελίδας Facebook, η οποία θα εξυπηρετούσε τους σκοπούς της εργασίας και θα είναι το μέσο για το «κατέβασμα» των δεδομένων προς την εφαρμογή. Η σελίδα που δημιουργήθηκε, το My Thesis Page, έχει έναν καθαρά δοκιμαστικό χαρακτήρα καθώς όλες οι αναρτήσεις, δεν εξυπηρετούσαν κάποιο ιδιαίτερο σκοπό-συμφέρον(οι αναρτήσεις ήταν αποφθέγματα και γνωμικά από το ίντερνετ) και ούτε είχε νόημα μία ραγδαία αύξηση της απήχησης. Αυτό που μας ενδιέφερε ήταν να υπάρχει ένας καλός αριθμός από ακόλουθους (>300 άτομα), ώστε κάθε φορά που γίνεται μια ανάρτηση να υπάρχει ένας ελάχιστος αριθμός ατόμων που θα δει την ανάρτηση (impressions) και θα αντιδράσει (reaction). Επίσης εξασφαλίζοντας τουλάχιστον >300 ακόλουθους έχουμε μια καλή βάση δεδομένων ώστε τα δημογραφικά στοιχεία που παίρνουμε από τους ακόλουθους της σελίδας να μπορούμε να τα αναλύσουμε ρεαλιστικά σαν προσομοίωση σελίδας στο Facebook από μία μικρομεσαία ελληνική επιχείρηση.

4.1.1 Facebook For Developers

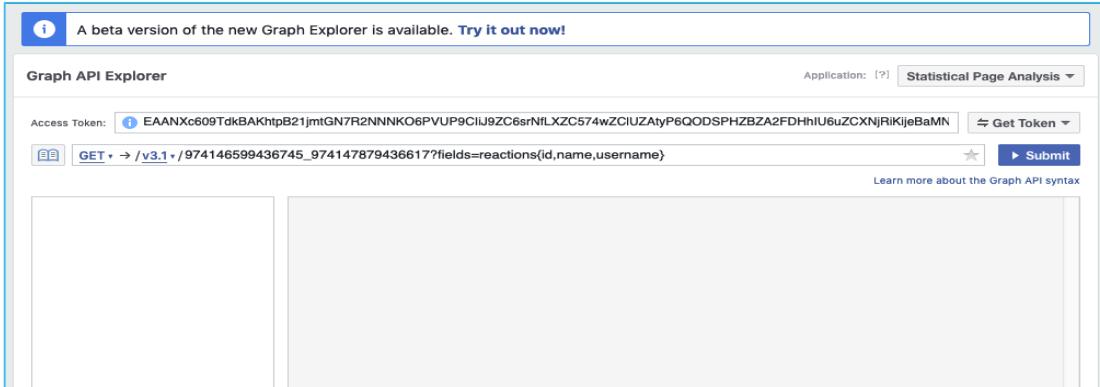
Πρώτο βήμα μετά την δημιουργία της σελίδας στο Facebook, ήταν να κάνουμε αίτηση στο Facebook για certified developer και να δημιουργήσουμε μια εφαρμογή μέσω της οποίας θα έχουμε πρόσβαση στα δεδομένα της σελίδας.



Εικόνα 4.1. To Dashboard στην αρχική σελίδα στο Facebook for Developers

Η διαδικασία αυτή διαρκεί σχεδόν 2 μέρες, καθώς ο αιτούμενος πρέπει να απαντήσει σε ορισμένες ερωτήσεις που θέτονται από το Facebook για τους σκοπούς της εφαρμογής και ποια δεδομένα πρόκειται να χρησιμοποιήσει. Εξαιτίας του νέου ευρωπαϊκού κανονισμού προστασίας δεδομένων το Facebook είναι υποχρεωμένο να κάνει αυστηρούς ελέγχους προς τους προγραμματιστές και να περιορίζει την πρόσβαση σε δεδομένα που δεν συνάδουν με τον σκοπό της εφαρμογής ή θεωρούνται ευαίσθητα σύμφωνα με τους κανονισμούς.

Αφού ολοκληρωθεί η διαδικασία, και έχουμε το token(κλειδί) της εφαρμογής, πηγαίνουμε στο Facebook Graph API Explorer.



Εικόνα 4.2. To Facebook Graph API Explorer.

Μέσω του Graph API Explorer παίρνουμε το access token για να μπορούμε να κάνουμε HTTP requests προς την σελίδα από την εφαρμογή που θα αναπτύξουμε. Βέβαια το Graph API Explorer προσφέρει και αυτό σαν εργαλείο τη δυνατότητα για HTTP request, πράγμα που το κάνει αρκετά χρήσιμο για γρήγορο έλεγχο των δεδομένων. Δηλαδή σε ποια έχουμε πρόσβαση και σε τη μορφή επιστρέφονται αυτά.

Μέσω του GET TOKEN και στη συνέχεια του GET Public Access Token εμφανίζεται λίστα με όλες τις δυνατές επιλογές, με δημόσια δεδομένα της σελίδας που θέλουμε να έχουμε πρόσβαση.

User Data Permissions	Events, Groups & Pages	Other
<input type="checkbox"/> email	<input type="checkbox"/> user_hometown	<input type="checkbox"/> user_posts
<input type="checkbox"/> user_age_range	<input type="checkbox"/> user_likes	<input type="checkbox"/> user_status
<input type="checkbox"/> user_birthday	<input type="checkbox"/> user_link	<input type="checkbox"/> user_tagged_places
<input type="checkbox"/> user_friends	<input type="checkbox"/> user_location	<input type="checkbox"/> user_videos
<input type="checkbox"/> user_gender	<input type="checkbox"/> user_photos	
Events, Groups & Pages		
<input type="checkbox"/> ads_management	<input type="checkbox"/> pages_manage_cta	<input type="checkbox"/> pages_show_list
<input type="checkbox"/> ads_read	<input type="checkbox"/> pages_manage_instant_articles	<input type="checkbox"/> publish_pages
<input type="checkbox"/> business_management	<input type="checkbox"/> pages.messaging	<input type="checkbox"/> publish_to_groups
<input type="checkbox"/> groups_access_member_info	<input type="checkbox"/> pages.messaging_phone_number	<input type="checkbox"/> read_page_mailboxes
<input type="checkbox"/> manage_pages	<input type="checkbox"/> pages.messaging_subscriptions	<input type="checkbox"/> user_events
Other		
<input type="checkbox"/> instagram_basic	<input type="checkbox"/> leads_retrieval	<input type="checkbox"/> read_insights
<input type="checkbox"/> instagram_manage_comments	<input type="checkbox"/> publish_video	<input type="checkbox"/> instagram_manage_insights
<input type="checkbox"/> read_audience_network_insights		

Εικόνα 4.3. Οι άδειες για τα δεδομένα που επιθυμούμε να πάρουμε

Αυτά που χρησιμοποιήσαμε ήταν όλα εκείνα της κατηγορίας User Data Permissions και από τα Other το read_insights, το οποίο είναι και το σημαντικότερο. Έχοντας το access token, μπορούμε

πλέον να κάνουμε HTTP requests στα πεδία που επιθυμούμε. Όλα τα request της σελίδας είναι με πεδία της κατηγορίας [Page/Insights](#).

Ανατρέχοντας σε αυτή την κατηγορία, βλέπει κανείς μια πλήρη λίστα με πεδία και μετρικές που μπορούν να χρησιμοποιηθούν για την εξόρυξη δεδομένων. Αλήθεια είναι πως από όσες μετρικές προσφέρονται, ορισμένες λειτουργούν σωστά. Αρκετές ενώ αναγράφεται πως είναι διαθέσιμες, το μόνο που επιστρέφουν μετά το request είναι το id της σελίδας, πράγμα που σημαίνει πως δεν έχουμε πρόσβαση. Αυτό μπορεί να οφείλεται είτε σε μια γενική αλλαγή πολιτικής προστασίας του Facebook στα διαθέσιμα δεδομένα του [Page/Insights](#) είτε σε στοχευμένο αποκλεισμό της εφαρμογής μας, επειδή δεν πληροί ορισμένα κριτήρια. Για παράδειγμα, ενώ μέχρι και τον Δεκέμβριο του 2018 υπήρχε πρόσβαση σε πεδία όπως: `page_content_activity_by_city_unique`, `page_content_activity_by_locale_unique`, `page_content_activity_by_age_gender_unique` από 1/1/2019 σταμάτησε η πρόσβαση, χωρίς να υπάρχει κάποια αλλαγή στην πολιτική προστασίας δεδομένων. Λόγω των γρήγορων και άμεσων αλλαγών θα πρέπει κάθε προγραμματιστής που θα ασχοληθεί με Facebook API Request, να δουλεύει με ένα μεγάλο σύνολο από διαφορετικά requests, ώστε να έχει εναλλακτικές λύσεις σε περίπτωση απαγόρευσης σε ορισμένα requests.

4.1.2. Δεδομένα και Μετρικές που Χρησιμοποιήθηκαν

Ακολουθεί η πλήρης λίστα των μετρικών που χρησιμοποιήθηκαν για την εξόρυξη δεδομένων :

Όνομα Μετρικής	Περιγραφή
<code>page_fans_locale</code>	Δεδομένα γλώσσας σχετικά με τους ακόλουθους της σελίδας βασισμένη στης ρυθμίσεις γλώσσας που έχουν επιλέξει.
<code>page_fans_city</code>	Δεδομένα τοποθεσίας, ταξινομημένα ανά πόλη των ακόλουθων της σελίδας.
<code>page_fans_country</code>	Δεδομένα τοποθεσίας, ταξινομημένα ανά χώρα των ακόλουθων της σελίδας.
<code>page_fans_gender_age</code>	Δημογραφικά δεδομένα σχετικά με τους ακόλουθους της σελίδας βασισμένα στην

	ηλικία και φύλο που έχουν ορίσει οι ακόλουθοι.
page_impressions_unique	Ο αριθμός των ατόμων που τους εμφανίστηκε οποιοδήποτε περιεχόμενο από την σελίδα στις οθόνες τους.
page_impressions_paid_unique*	Ο αριθμός των ατόμων που τους εμφανίστηκε οποιοδήποτε περιεχόμενο από την σελίδα στις οθόνες τους, μέσω πληρωμένου μέσου π.χ. διαφήμισης.
page_impressions_organic_unique	Ο αριθμός των ατόμων που τους εμφανίστηκε οποιοδήποτε περιεχόμενο από την σελίδα στις οθόνες τους, μέσω μη πληρωμένου μέσου.
page_fan_adds_by_paid_non_paid_unique	Ο αριθμός των ατόμων που πάτησαν “Like” στη σελίδα, χωρισμένος σε πληρωμένα και μη μέσα.
page_views_total	Ο αριθμός των προβολών της σελίδας από συνδεδεμένα και μη άτομα.
page_fans	Ο Συνολικός αριθμός των ατόμων που τους «Αρέσει» η σελίδα.
page_posts_impressions	Ο αριθμός των προβολών οποιασδήποτε ανάρτησης της σελίδας
page_post_engagements	Ο αριθμός των διαδράσεων από ακόλουθους σε οποιασδήποτε ανάρτηση (Likes, Comments, Shares).
page_consumptions	Ο αριθμός των κλικ που είχε οποιοδήποτε περιεχόμενο της σελίδας.

page_posts_impressions_paid	Ο αριθμός των εμφανίσεων οποιασδήποτε ανάρτησης στην οθόνη κάποιου ατόμου μέσω διαφήμισης.
page_posts_impressions_organic	Ο αριθμός των εμφανίσεων οποιασδήποτε ανάρτησης στην οθόνη κάποιου ατόμου μέσω μη πληρωμένου μέσου.
page_posts_impressions_viral	Ο αριθμός των εμφανίσεων οποιασδήποτε ανάρτησης στην οθόνη κάποιου ατόμου συνοδευόμενη με κάποια κοινωνική πληροφορία. (Όταν κάποιος φύλος μου του αρέσει μία ανάρτηση και αυτή εμφανίζεται και σε μένα)
post_impressions	Ο αριθμός των εμφανίσεων της συγκεκριμένης ανάρτησης στις οθόνες των ακολούθων.
post_impressions_paid_unique	Ο αριθμός των ατόμων που τους εμφανίστηκε η συγκεκριμένη ανάρτηση στις οθόνες τους, μέσω πληρωμένου μέσου π.χ. διαφήμισης.
post_impressions_organic_unique	Ο αριθμός των ατόμων που τους εμφανίστηκε η συγκεκριμένη ανάρτηση στις οθόνες τους, μέσω μη πληρωμένου μέσου.
post_impressions_fan_unique	Ο αριθμός των ακολούθων που είδαν την συγκεκριμένη ανάρτηση.
post_impressions_fan_unique	Ο αριθμός των ακολούθων που είδαν την συγκεκριμένη ανάρτηση, μέσω πληρωμένου μέσου π.χ. διαφήμισης.

post_engaged_users	Ο αριθμός των ατόμων που έκαναν κλικ οπουδήποτε πάνω στην ανάρτηση.
post_impressions_viral_unique	Ο αριθμός των ατόμων που τους εμφανίστηκε η συγκεκριμένη ανάρτηση συνοδευμένη με κάποια κοινωνική πληροφορία.

Χρησιμοποιώντας τις παραπάνω μετρικές κάνουμε requests της μορφής:

```
page_fans_city =
graph.get_object(id='974146599436745', fields='insights.metric(page_fans_city)')
```

Αποθηκεύουμε τα δεδομένα που επιστρέφει το request στην αντίστοιχη μεταβλητή. Ως παραμέτρους στο request έχουμε το id, το οποίο είναι είτε το id της σελίδας ή κάποιας ανάρτησης και την εκάστοτε μετρική που θα χρησιμοποιήσουμε.

Τα δεδομένα που επιστρέφονται από το request είναι σε μορφή JSON dictionary με το εξής pagination :

```
(dict){      insights(dict){      data(list)[      (dict){      name:"_", period:"_", values: (list)[      (dict){      value:"_", end_time:"_"}]}]}}
```

Έτσι αν θέλουμε να πάρουμε π.χ. την τιμή των page_fans και την ημερομηνία που αντιστοιχεί στην τιμή αυτή θα κάνουμε:

```
page_fans_value = page_fans["insights"]["data"][0]['values'][1]['value']
date = page_fans["insights"]["data"][0]['values'][1]['end_time']
```

Αξίζει να σημειωθεί πως τα δεδομένα που επιστρέφονται έχουν καθυστέρηση 2 ημέρων. Δηλαδή οποιαδήποτε μέρα και ώρα γίνεται το request, η τιμή που θα μου επιστραφεί είναι η τιμή που έχει η σελίδα 2 μέρες πριν στις 7:00 το πρωί. Οπότε δεν έχει νόημα να γίνουν περισσότερα από ένα request στην ίδια μετρική μέσα σε μία ημέρα.

```

"insights": {
    "data": [
        {
            "name": "page_fans",
            "period": "day",
            "values": [
                {
                    "value": 473,
                    "end_time": "2019-04-08T07:00:00+0000"
                },
                {
                    "value": 472,
                    "end_time": "2019-04-09T07:00:00+0000"
                }
            ],
            "title": "Lifetime Total Likes",
            "description": "Lifetime: The total number of people who have liked your page at least once since it was created.",
            "id": "974146599436745/insights/page_fans/day"
        }
    ]
}

```

Εικόνα 4.4.Το JSON Dictionary που επιστρέφει το request.

4.1.3. Δημογραφικά Δεδομένα

Τα σημαντικότερα δεδομένα που παίρνουμε από τα requests και αυτά που αναλύουμε περισσότερο είναι τα δημογραφικά δεδομένα που αφορούν τους ακόλουθους της σελίδας. Ηλικία, Φύλο, Πόλη, Χώρα, Γλώσσα είναι τα δεδομένα που έχουμε πρόσβαση. Οι εκάστοτε τιμές αυτών αφορούν το σύνολο των ακολούθων της σελίδας και μόνο.

- **Ηλικία και Φύλλο**

Τα δεδομένα που επιστρέφονται για την ηλικία και το φύλο των ακόλουθων της σελίδας είναι χωρισμένα σε 14 ομάδες-κατηγορίες:

F.13-17	F.18-24	F.25-34	F.35-44	F.45-54	F.55-64	F.65+
M.13-17	M.18-24	M.25-34	M.35-44	M.45-54	M.55-64	M.65+

Με F. αναγράφονται οι γυναίκες και με M. οι άνδρες. Σε καθένα από τα δύο φύλα υπάρχουν 7 ηλικιακές ομάδες. Η ηλικία και το φύλο των ακολούθων προκύπτει βάση της ημερομηνίας γέννησης και το φύλο που έχουν δηλώσει κατά την εγγραφή τους στο Facebook.

Κάθε φορά που γίνεται request επιστρέφεται ένα dictionary με κλειδιά (keys) τις παραπάνω κατηγορίες και τιμή (value) το πλήθος των ακολούθων που ανήκουν σε αυτή την ομάδα.

- **Πόλη**

Σε αντίθεση με τις ηλικιακές ομάδες, όταν γίνεται request για τις πόλεις που μένουν οι ακόλουθοι της σελίδας, επιστρέφεται σχεδόν διαφορετικό dictionary κάθε φορά. Κλειδιά σε αυτό το dictionary είναι οι πόλεις και τιμή το πλήθος των ατόμων που μένουν σε αυτή τη πόλη. Οι ονομασίες των πόλεων έχουν τη μορφή Πόλη, Διαμέρισμα, Χώρα. Π.χ. η Θεσσαλονίκη αναγράφεται ως Thessaloniki, Central Macedonia, Greece. Αυτό είναι ένα αρκετά βοηθητικό στοιχείο καθώς γνωρίζοντας το γεωγραφικό διαμέρισμα κάθε πόλης μπορούμε να κάνουμε ομαδοποίηση των δεδομένων των πόλεων και μια ευρύτερη ανάλυση στα γεωγραφικά διαμερίσματα.

- Χώρα**

Όπως και στις πόλεις, έτσι και στις χώρες μας επιστρέφεται ένα dictionary με τις χώρες στις οποιες μένουν οι ακόλουθοι της σελίδας. Στο dictionary οι χώρες αναγράφονται με τον διεθνή κωδικό 2 γραμμάτων. Π.χ. η Ελλάδα αναγράφεται ως GR, η Κύπρος ως CY και ούτω καθεξής.

- Γλώσσα**

Στο συγκεκριμένο dictionary εμφανίζονται οι γλώσσες που μιλούν οι ακόλουθοι της σελίδας. Ουσιαστικά είναι η γλώσσα που έχουν επιλέξει οι ακόλουθοι στις ρυθμίσεις του Facebook.

4.1.4. Αποθήκευση Δημογραφικών Δεδομένων

Όλα αυτά τα δεδομένα που λαμβάνουμε από τα JSON dictionaries που μας επιστρέφουν τα request, αποθηκεύονται σε αρχεία excel. Ο λόγος που επέλεξα τα αρχεία excel έναντι σε csv ή μια κλασσική SQL βάση, είναι πως εύκολα ο χρήστης θα μπορεί από παντού σχεδόν να ανοίξει ένα excel, να δει και να ελέγχει όλα τα δεδομένα της σελίδας του. Επίσης τα αρχεία excel είναι εύκολο να διαχειριστούν και να αναλυθούν μέσω των βιβλιοθηκών Pandas και NumPy που χρησιμοποιήθηκαν κατά κόρων στην εφαρμογή μέσα.

A	B	C	D
1 Date	F.13-17	F.18-24	F.25-34
2 01/02/2019	2	42	49
3 02/02/2019	2	45	49
4 03/02/2019	2	45	49
5 04/02/2019	2	42	49
6 05/02/2019	2	45	49
7 06/02/2019	2	51	49
8 07/02/2019	2	51	49
9 08/02/2019	2	51	49

A	B	C	D
1 Date	Maputo, M	Bratislava, I	Agrínio, W
2 01/02/2019	2	2	4
3 02/02/2019	2	2	4
4 03/02/2019	2	2	4
5 04/02/2019	2	2	4
6 05/02/2019	2	2	4
7 06/02/2019	2	2	4
8 07/02/2019	2	2	4
9 08/02/2019	2	2	4

Εικόνα 4.5 Αριστερά το αρχείο που αποθηκεύονται ηλικίες και δεξιά οι πόλεις των ακόλουθων.

Συνολικά υπάρχουν 10 αρχεία excel στα οποία αποθηκεύονται τα δεδομένα της σελίδας και από τα οποία αντλεί η εφαρμογή τα δεδομένα για την ανάλυση

Ακολουθεί λίστα με τα ονόματα των αρχείων και την αναλυτική περιγραφή τους:

Αρχείο	Περιγραφή
Age-Gender.xlsx	Φύλλο και ηλικίες των ακολούθων.
City.xlsx	Πόλεις που μένουν οι ακόλουθοι.
Country.xlsx	Χώρες στις οποίες μένουν οι ακόλουθοι.
Locale.xlsx	Γλώσσες που μιλάνε οι ακόλουθοι.
Page-Info.xlsx	Βασικές Πληροφορίες για τη σελίδα.
Page-Post.xlsx	Βασικές Πληροφορίες για τις αναρτήσεις της σελίδας.
Post-Info.xlsx	Οι τελευταία ενημερωμένες πληροφορίες για κάθε συγκεκριμένη ανάρτηση της σελίδας.
Post-Info-Continuously.xlsx	Πληροφορίες για κάθε συγκεκριμένη ανάρτηση της σελίδας σε πάροδο ημερών.
Region-Dataframe.xlsx	Ο τελευταία ενημερωμένος κατάλογος με τα top-7 γεωγραφικά διαμερίσματα που μένουν οι ακόλουθοι της σελίδας.
Region-Dataframe-Continuously.xlsx	Κατάλογος με τα top-7 γεωγραφικά διαμερίσματα που μένουν οι ακόλουθοι της σελίδας σε πάροδο ημερών.

4.2 Υλοποίηση Εφαρμογής και Ανάπτυξη Κώδικα

Η εφαρμογή που αναπτύχθηκε για την εξόρυξη δεδομένων από σελίδα Facebook, ανάλυση αυτών και εξαγωγή συμπερασμάτων και γνώσης, είναι γραμμένη εξ ολοκλήρου σε γλώσσα Python.

Η έκδοση που χρησιμοποιήθηκε ήταν η 3.6 από Anacoda Distribution σε περιβάλλον Visual Code Studio. Συνολικά η εφαρμογή είναι αναπτυγμένη σε 10 αρχεία συνολικής έκτασης ~3.500 σειρών κώδικα.

4.2.1. Βιβλιοθήκες και Packages

Κατά την ανάπτυξη του κώδικα χρησιμοποιήθηκαν και ορισμένες βιβλιοθήκες-packages που διευκόλυναν αρκετά την υλοποίηση της εφαρμογής. Όλες εγκαταστάθηκαν μέσω της εντολής pip install package σε terminal και χρησιμοποιήθηκαν μέσω import μέσα στον κώδικα. Αυτές είναι :

Json	Facebook	Xlsxwriter	Datetime
Openpyxl	Pandas	NumPy	Unicode
Matplotlib	Tabulate	Plotly	Dash
Pycountry	Dash_core_components	Dash_html_components	Dash.dependencies

4.2.2. Αρχεία Python

Όπως προαναφέρθηκε η εφαρμογή αποτελείται από 10 αρχεία γλώσσας Python. Σε ορισμένα γίνεται εξόρυξη δεδομένων από το Graph API του Facebook και αποθήκευση σε excel και σε άλλα γίνεται ανάλυση και παρουσίαση των δεδομένων.

- Demographics.py

Σε αυτό το αρχείο, γίνεται εξόρυξη των δημογραφικών δεδομένων της σελίδας (Ηλικίες, Πόλεις, Χώρες, Γλώσσες) και αποθήκευση αυτών στα αντίστοιχα excel αρχεία. Σε κάθε περίπτωση υπάρχουν δυο ενδεχόμενα για τα οποία γίνεται έλεγχος: 1) αν το excel αρχείο (age-gender.xlsx, city.xlsx, country.xlsx, locale.xlsx) υπάρχει, ώστε να γίνει update στις στήλες και 2) αν το excel αρχείο δεν υπάρχει, ώστε να δημιουργηθεί και να δομηθούν οι στήλες εξ αρχής. Παράλληλα γίνεται έλεγχος για κενά κελιά, στα οποία εισάγεται 0. Ένα άδειο κελί προκύπτει όταν σε προηγούμενο request είχαμε τιμή για μια ηλικία-πόλη-χώρα-γλώσσα και πλέον επειδή δεν υπάρχει πλέον κάποιος ακόλουθος αυτής της κατηγορίας (unlike) δεν επιστρέφεται τίποτα. Είναι αναγκαίο κανένα κελί σε όλα τα excel αρχεία να μην είναι άδειο για να γίνει σωστά η ανάλυση των δεδομένων.

- Fetch_regions.py

Εδώ ο αλγόριθμος παίρνει τα ονόματα των πόλεων και τις τιμές τους από το city.xlsx αρχείο και τα ομαδοποιεί ανάλογα με τα γεωγραφικά διαμερίσματα. Π.χ. οι πόλεις Thessaloniki,

Central Macedonia, Greece, Thérmi, Central Macedonia, Greece και Kilkís, Central Macedonia, Greece θα ομαδοποιηθούν ως Central Macedonia.

Η τιμή που έχει το κάθε γεωγραφικό διαμέρισμα είναι το άθροισμα των τιμών των πόλεων που την απαρτίζουν. Από το σύνολο που δημιουργείται κρατάμε τα top 7 γεωγραφικά διαμερίσματα, καθώς τα υπόλοιπα έχουν αρκετά μικρές τιμές χωρίς ιδιαίτερο νόημα για ανάλυση.

Βέβαια αυτό μπορεί να αλλάξει εύκολα αν οι ακόλουθοι αυξηθούν και οι απαρτήσεις της σελίδας είναι μεγαλύτερες. Από εδώ παράγονται τα δύο αρχεία excel που αποθηκεύουμε τις τιμές των γεωγραφικών διαμερισμάτων, το regionDF.xlsx που κρατά τις μέσες τιμές των 7 μεγαλύτερων διαμερισμάτων και το regionDF_continuously.xlsx που κρατά συνεχόμενα για κάθε μέρα τις τιμές των διαμερισμάτων.

- [Page_info.py](#)

Εδώ λαμβάνουμε τα dictionaries που αφορούν τις γενικές πληροφορίες της σελίδας: Impressions, Impressions paid, Impressions organic, Fan adds, Page views, Fan adds by paid – non paid. Πάλι γίνεται έλεγχος αν το αρχείο excel υπάρχει ή όχι, ώστε να προστεθούν απλά οι νέες τιμές ή να γίνει εκ νέου διαμόρφωση αρχείου. Τα αποτελέσματα που προκύπτουν για κάθε μέρα ξεχωριστά αποθηκεύονται στο page-info.xlsx.

- [Page_post_info.py](#)

Στο page_post_info έχουμε να κάνουμε με πληροφορίες που σχετίζονται με τις συνολικές αναρτήσεις της σελίδας: Post impressions, Post impressions paid, Post impressions organic, Post impressions viral, Post engagement, Post consumptions. “Ολες αυτές οι τιμές μαζί με την ημερομηνία που αντιστοιχούν αποθηκεύονται στο page-post.xlsx.

- [Post.py](#)

Σε αυτό το αρχείο παίρνουμε από το Facebook μια λίστα με τις αναρτήσεις που έχουν γίνει στη σελίδα και τις πληροφορίες που συνοδεύονται για την κάθε ανάρτηση ξεχωριστά. Οι πληροφορίες αυτές είναι : ID ανάρτησης, Ημερομηνία δημιουργίας, Μήνυμα, Impressions, Impressions paid, Impressions organic, Impressions fans, Impressions fans paid, Engaged users, Impressions viral, Total reactions.

Στη συνέχεια οι πληροφορίες αυτές αποθηκεύονται στα αρχεία excel post-info και στο post-info-continuously . Στο πρώτο διατηρούμε μόνο τις πιο πρόσφατες πληροφορίες για τις αναρτήσεις κάθε φορά ενώ στο δεύτερο διατηρούμε τις πληροφορίες των αναρτήσεων για κάθε μέρα ξεχωριστά.

- Show_info_stats.py

Εδώ αναλύονται τα δεδομένα από όλα τα αρχεία excel και παρουσιάζονται στατιστικά στοιχεία για τη σελίδα. Ακολουθούν ορισμένα από τα αποτελέσματα που εξάγονται :

		start	end	mean	growth in %	Largest Value	\
Page View Totals		0	81	39.705882	inf	265	
Page Fans		348	472	367.215686	3.563218e+01	473	
Page Fan Adds Paid		0	0	0.019608	Nan	1	
Page Fan Adds Non Paid		0	0	2.470588	Nan	73	
Page Impressions		1	1243	673.215686	1.242000e+05	3893	
Page Impressions Paid		0	0	259.196078	Nan	2001	
Page Impressions Organic		1	1243	428.039216	1.242000e+05	2738	

		start	end	mean	growth in %
Page Post Impressions		3	2094	898.714286	69700.0
Page Post Engagements		2	175	107.785714	8650.0
Page Consumptios		2	87	51.232143	4250.0
Page Post Impressions Paid		0	0	250.821429	Nan
Page Post Impressions Organic		3	2094	647.535714	69700.0
Page Post Impressions Viral		0	1965	520.678571	inf

Εικόνα 4.6. Στατιστικά στοιχεία για την σελίδα και τις αναρτήσεις της σελίδας

These are the top 4 Ages stats of the page
percent
M.18-24 21.277773
F.25-34 14.358809
F.18-24 14.137550
F.45-54 10.726468

Εικόνα 4.7. Στατιστικά στοιχεία για τις ηλικίες των ακόλουθων της σελίδας.

- elstat_correlation.py

Στο αρχείο αυτό γίνεται παρουσίαση των στατιστικών δεδομένων της ΕΛΣΤΑΤ αλλά και συσχέτιση αυτών με τα δεδομένα της σελίδας μας. Τα αρχεία που επιλέχθηκαν από την ΕΛΣΤΑΤ έχουν να κάνουν με το επίπεδο εκπαίδευσης στην Ελλάδα, την κατάσταση ασχολίας των Ελλήνων και την οικογενειακή κατάσταση. Όλα αυτά θα αναλυθούν ειδικότερα παρακάτω. Η συσχέτιση των δεδομένων γίνεται σε επίπεδο γεωγραφικών διαμερισμάτων στην Ελλάδα. Ακολουθούν ορισμένα από τα αποτελέσματα που εξάγονται:

Central Macedonia, Greece
Gender Total Higher Education Middle Education Lower Education
4 Male 912693 137009--[15.01%] 293707--[32.18%] 359928--[39.44%]
5 Female 969415 156144--[16.11%] 278246--[28.7%] 417474--[43.06%]
Facebook Page Correlation:
Higher Edu: Male=[14.39] Female=[15.44]
Middle Edu: Male=[30.85] Female=[27.51]
Lower Edu: Male=[37.81] Female=[41.28]

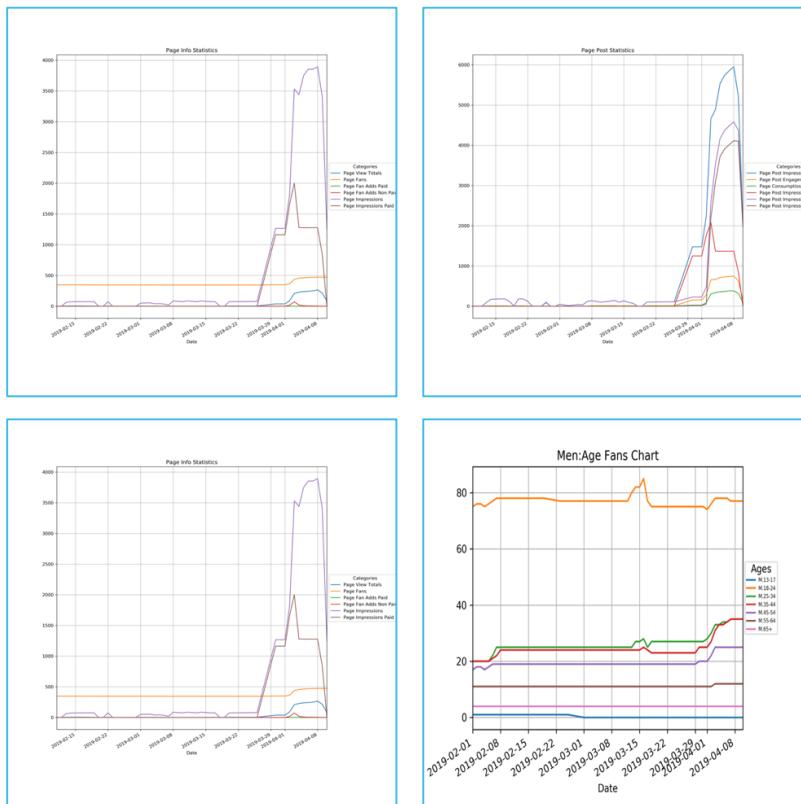
AGE AND OCCUPATIONAL STATS					
Top 4 Age-Gender Groups:					
	Age	Total	Active %	Employed\n %	Unemployed
8	M.18-24	53.77	31.66		22.11
2	F.25-34	73.41	55.79		17.62
1	F.18-24	46.08	24.90		21.17
4	F.45-54	56.31	49.19		7.12

Facebook Page Correlation:					
Fans of Both Genders Non Married in FB Page : [36.92]					
Fans of Both Genders Married in FB Page : [48.27]					
Fans of Both Genders Widower in FB Page : [7.37]					
Fans of Both Genders Divorced in FB Page : [2.64]					
Fans of Both Genders Civil Partnership in FB Page : [0.01]					
Fans of Both Genders Separated in FB Page : [0.63]					
Fans of Both Genders Widower From Civil Partnership in FB Page : [0.]					
Fans of Both Genders Separated From Civil Partnership in FB Page : [0.]					

Εικόνα 4.8. Τα αποτελέσματα που εξάγονται από την συσχέτιση τα δεδομένα της ΕΛΣΤΑΤ.

- Charts.py

Εδώ γίνεται χρήση της βιβλιοθήκης matplotlib για να παραχθούν γραφήματα σε μορφή εικόνων png. Βέβαια η χρήση τους δεν εξυπηρετεί ιδιαίτερα καθώς μετέπειτα αναπτύχθηκε διαδραστικό dashboard που παρουσιάζει τα δεδομένα καλύτερα. Μερικά από τα γραφήματα που παράγονται είναι τα εξής:



Εικόνα 4.9. Γραφήματα με πληροφορίες της σελίδας μέσω του matplotlib

- Plotly_elstat.py

Εδώ με τη χρήση του Dash και του Plotly φτιάχνουμε μια εφαρμογή Flask η οποία τρέχει σε server είτε σε local host είτε online αν γίνει deployment. Στην εφαρμογή αυτή παρουσιάζονται διαδραστικά τα στοιχεία της ΕΛΣΤΑΤ που χρησιμοποιούμε, δηλαδή 1) οικογενειακή κατάσταση, 2) επίπεδο εκπαίδευσης, 3) κατάσταση ασχολίας. Τα δεδομένα μετατρέπονται αρχικά σε pandas dataframes και στη συνέχεια αναλύουμε κάθε φορά τα τμήματα που θέλουμε. Η διαδραστικότητα και η δυνατότητα live-update γίνεται μέσω των callbacks στις functions του αρχείου. Σαν παράμετροι δίνονται οι επιλογές από το dropdown menu και το submit button.

```
Running on http://127.0.0.1:8050/
Debugger PIN: 523-962-152
* Serving Flask app "plotly_elstat" (lazy loading)
* Environment: production
| WARNING: Do not use the development server in a production environment.
| Use a production WSGI server instead.
* Debug mode: on
Running on http://127.0.0.1:8050/
Debugger PIN: 643-335-737
```

Εικόνα 4.10. Η εφαρμογή Flask με τα γραφήματα της ΕΛΣΤΑΤ

Η εφαρμογή αυτή έχει ανέβει σε online server και είναι διαθέσιμη προς ελεύθερη χρήση στον σύνδεσμο <https://andreou-elstat.herokuapp.com/>.

```
@app.callback(
    Output('occupation-graph', 'figure'),
    [Input('submit-button-region', 'n_clicks')],
    [State('my_ticker_symbol_region', 'value')])
def update_occupation_graph(n_clicks, posts_ticker_cont):

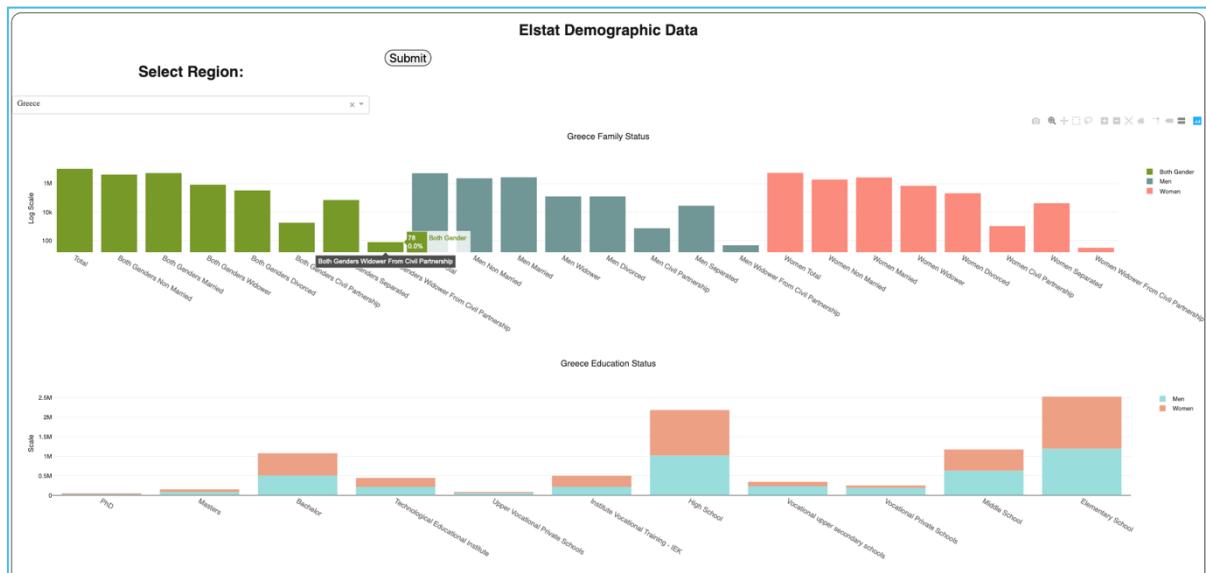
    age_dict = {}
    percent_age_dict={}
    for age in agelist:
        age_dict.update({age:[()]})
        percent_age_dict.update({age:[()]})

    for item in occupation_labels:
        for age in agelist:
            age_dict[age].append(df_occupation[(df_occupation['Region']==posts_ticker_cont)&(df_occupation['Age Group']==item)])
            temp = round(df_occupation[(df_occupation['Region']==posts_ticker_cont)&(df_occupation['Age Group']==item)].shape[0]/n_clicks)
            percent_age_dict[age].append(str(temp)+'%')
```

Εικόνα 4.11 Τα callbacks στις μεθόδους του plotly_elstat.py

Στην εφαρμογή αυτή κατηγοριοποιούμε τον ελληνικό πληθυσμό ή τον πληθυσμό ενός γεωγραφικού διαμερίσματος στις δύο πρώτες κατηγορίες δεδομένων ανάλογα με το φύλο και στην τελευταία

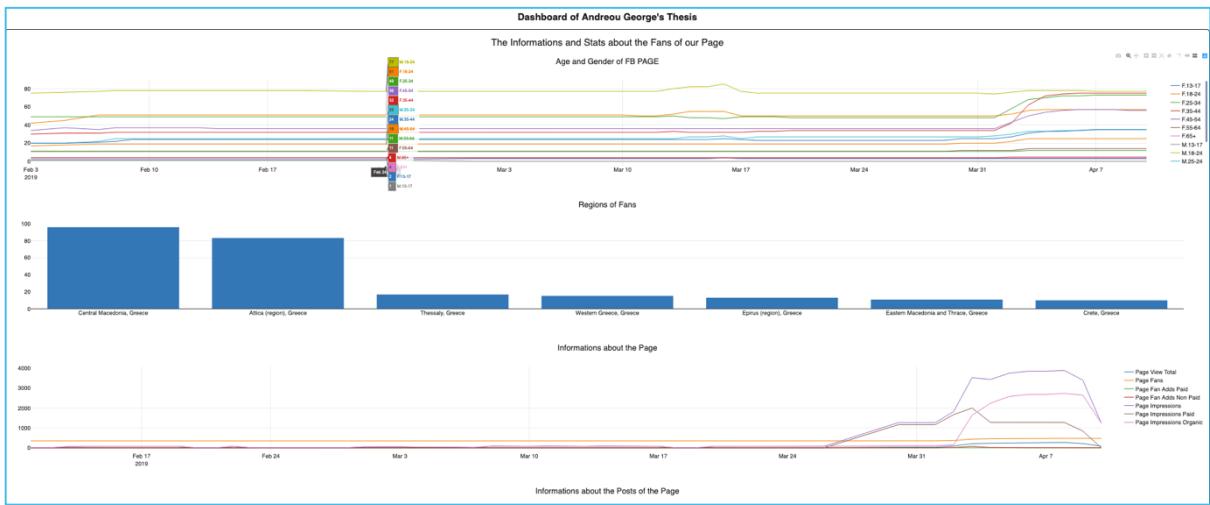
(κατάσταση ασχολίας) ανάλογα με φύλο και ηλικία. Ο χρήστης κάνοντας hover πάνω από τη κάθε μπάρα του γραφήματος, του εμφανίζεται ο αριθμός των ατόμων που ανήκουν σε αυτή τη κατηγορία και το ποσοστό που αντιπροσωπεύουν. Ως προεπιλογή εμφανίζονται τα δημογραφικά δεδομένα που αφορούν όλη την Ελλάδα ευρύτερα. Ο χρήστης μπορεί να επιλέξει από το dropdown μενού το γεωγραφικό διαμέρισμα που θέλει να εξετάσει και στη συνέχεια για να εμφανιστούν τα δεδομένα χρειάζεται μόνο να πατήσει submit .



Εικόνα 4.12 Τα δημογραφικά δεδομένα της ΕΛΣΤΑΤ στην εφαρμογή.

- `Plotly_charts.py`

Σε αυτό το αρχείο πάλι με τη βοήθεια του Dash και του Plotly φτιάχνουμε μια εφαρμογή Flask για να εμφανίσουμε μέσω διαδραστικών διαγραμμάτων τις πληροφορίες της σελίδας. Εδώ γίνεται χρήση όλων των αρχείων excel που αποθηκεύαμε τιμές προηγουμένως ώστε να δημιουργήσουμε μία συνολική απεικόνιση των πληροφοριών, των δημογραφικών στοιχείων της σελίδας και της συσχέτισης και αλληλεπίδρασης όλων αυτών. Στο Dashboard εμφανίζονται γραφήματα για τις ηλικίες-φύλο, τα γεωγραφικά διαμερίσματα, τις πληροφορίες της σελίδας, τις πληροφορίες των αναρτήσεων, τις πόλεις και χώρες των ακολούθων και τις πληροφορίες κάθε ανάρτησης ξεχωριστά. Στο τέλος εμφανίζεται μια σειρά γραφημάτων που αφορούν το κομμάτι του spending strategy το οποίο θα αναλυθεί λεπτομερώς στο επόμενο κεφάλαιο.



Εικόνα 4.13 Το dashboard της σελίδας με τα 3 πρώτα γραφήματα.

4.3 Δεδομένα ΕΛΣΤΑΤ

Η Ελληνική Στατιστική Αρχή (ΕΛΣΤΑΤ) είναι η εθνική στατιστική υπηρεσία της Ελλάδας με σκοπό τη συστηματική παραγωγή επίσημων στατιστικών, καθώς και τη διενέργεια στατιστικών ερευνών. Οι έρευνες αυτές 1) αφορούν όλους τους τομείς της δραστηριότητας τόσο του δημοσίου όσο και του ιδιωτικού τομέα, 2) υποστηρίζουν τη διαδικασία λήψης αποφάσεων, χάραξης και αξιολόγησης πολιτικών της Κυβέρνησης και των φορέων του Δημοσίου (δείκτες αξιολόγησης), 3) υποβάλλονται σε διεθνείς φορείς σύμφωνα με τις υποχρεώσεις της χώρας και 4) αφορούν στο γενικό κοινό ή κατηγορίες χρηστών στατιστικών στοιχείων στο εσωτερικό ή στο εξωτερικό. Η Ελληνική Στατιστική Αρχή συλλέγει στατιστικά στοιχεία που αφορούν τον πληθυσμό (είναι υπεύθυνη για την διεξαγωγή της απογραφής πληθυσμού κάθε 10 χρόνια), την υγεία και την κοινωνική ασφάλιση, την απασχόληση και την ανεργία, την εκπαίδευση κλπ. Χρήση των στατιστικών στοιχείων που συλλέγει κάνουν τόσο το ελληνικό κράτος, όσο και διεθνείς οργανισμοί, αλλά και επιχειρήσεις, επιστήμονες και απλοί πολίτες. [38]

4.3.1 Δεδομένα που Χρησιμοποιήθηκαν

Τα δεδομένα που κάναμε χρήση ανήκουν στη κατηγορία των Δημογραφικών και Οικονομικών χαρακτηριστικών του 2011. Πιο συγκεκριμένα τα αρχεία που χρησιμοποιήσαμε ήταν τα εξής :

- Α02. Μόνιμος Πληθυσμός κατά ηλικία, φύλο και οικογενειακή κατάσταση Σύνολο χώρας, Περιφέρειες (Δημογραφικά Χαρακτηριστικά)
- Α01. Μόνιμος Πληθυσμός, κατά φύλο, ομάδες ηλικιών και κατάσταση ασχολίας. Σύνολο Χώρας, Περιφέρειες (Οικονομικά Χαρακτηριστικά)

- Α03. Μόνιμος Πληθυσμός, κατά φύλο, ομάδες ηλικιών και επίπεδο εκπαίδευσης Σύνολο χώρας, Περιφέρειες (Δημογραφικά Χαρακτηριστικά)

Όλα αυτά τα αρχεία είναι σε μορφή excel. Βέβαια ο τρόπος δόμησης του αρχείου και οι κατηγορίες-ομάδες των δεδομένων δεν είναι εύχρηστος για την δικιά μας περίπτωση. Έτσι τροποποιήσαμε τα αρχεία και τα φέραμε σε μορφή που να ταιριάζει με τα δεδομένα που παίρνουμε από το Facebook.

Για παράδειγμα στο αρχείο A02 που αφορά την οικογενειακή κατάσταση των Ελλήνων η μορφή των δεδομένων είναι η εξής:

Εικόνα 4.14 Το αρχείο της ΕΛΣΤΑΤ για την οικογενειακή κατάσταση

Στα αρχεία της ΕΛΣΤΑΤ αναγράφονται πληροφορίες και κατηγορίες όπως γεωγραφικό επίπεδο, γεωγραφικός κώδικας καθώς και αναλυτικά τα στοιχεία των πόλεων κάθε γεωγραφικού διαμερίσματος. Επίσης οι ηλικιακές ομάδες που χρησιμοποιούνται για την κατηγοριοποίηση του πληθυσμού είναι περισσότερες από αυτές που χρειαζόμαστε και δουλεύουμε με τα δεδομένα του Facebook. Μια τόσο λεπτομερής ανάλυση σε βαθμό πόλεων και περισσότερων ηλικιακών ομάδων δεν θα ωφελούσε στην παρούσα στιγμή. Η σελίδα που έχουμε φτιάξει στο facebook δεν έχει τόσο πολλούς ακόλουθους στη σελίδα ώστε να χρειάζεται τόσο λεπτομερής ανάλυση.

Για αυτό το λόγο ήταν ανάγκη να τροποποιήσουμε τα αρχεία της ΕΛΣΤΑΤ. Οι ενέργειες που έγιναν ήταν οι εξής :

- Ομαδοποίηση των δεδομένων και άθροιση των τιμών σε επίπεδο γεωγραφικού διαμερίσματος.
 - Μετατροπή του ηλικιακού εύρους σύμφωνα με την μορφή του Facebook
 - Προσθήκη νέας στήλης με ηλικία και φύλο π.χ. (F.35-44)
 - Άρση της ομαδοποίησης(Ungroup) στις στήλες, ώστε κάθε κατηγορία να έχει τη δικιά της στήλη στο αρχείο.

- Μετάφραση των ονομάτων των στηλών-κατηγοριών στα αγγλικά, ώστε να γίνεται εύκολα αναζήτηση μέσα στα dataframes.

	A	B	C	D	E	F	G	H	I	J	K	L
1	Region	Total	Both Genders Non Married	Both Genders Married	Both Genders Widower	Both Genders Divorced	Both Genders Partnership	Both Genders Separated	Both Genders Widower From Civil Partnership	Both Genders Separated From Civil Partnership	Men Total	Men Non Married
2	Eastern Macedonia and Thrace, Greece	608.182	217.990	321.896	51.787	13.544	74	2.858	4	29	299.643	122.901
3	Central Macedonia, Greece	1.862.108	724.988	947.919	144.767	51.831	178	12.368	4	53	912.693	391.697
4	Western Macedonia, Greece	283.889	104.797	149.541	23.140	5.106	12	1.984	1	8	141.779	59.621
5	Epirus (region), Greece	336.856	124.469	176.740	28.037	6.296	21	1.281	0	12	165.775	68.309
6	Thessaly, Greece	732.762	270.507	383.692	60.301	15.041	104	3.076	6	35	362.194	151.890
7	Central Greece (region), Greece	547.390	202.317	283.985	45.976	12.562	87	2.433	9	21	277.475	118.341
8	Greece (island Islands (region)), Greece	207.855	76.680	106.226	16.326	7.086	67	1.466	1	3	102.400	42.387
9	Western Greece, Greece	679.796	279.597	328.478	52.750	15.612	115	3.222	2	20	339.310	157.198
10	Peloponnese (region), Greece	577.903	214.092	297.807	48.377	14.822	160	2.817	2	26	291.777	124.723
11	Attica (region), Greece	3.828.434	1.563.975	1.801.598	272.290	157.894	673	32.054	41	89	1.845.663	840.143
12	Northern Aegean, Greece	199.231	73.921	101.686	17.160	5.341	24	1.080	0	10	99.984	43.402
13	Southern Aegean, Greece	309.015	122.855	155.775	18.050	9.340	84	2.196	5	10	155.865	69.395
14	Crete, Greece	623.065	251.188	309.710	41.488	16.701	166	3.783	3	16	308.665	138.599
15	Greece	10.816.286	4.227.476	5.364.763	820.449	331.686	1.765	69.737	78	332	5.303.223	2.329.207

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	
1	Region	Gender	Age	Total	Total Active	Employed	Total Unemployed	Employed	Unemployed	Youth	Total Non Active Students	Retired	Rentier	Housekeeping	Rest
2	Macdonia and Thrace, Greece	Total Gender	Total Age	608.182	234.579	187.306	47.273	33.000	14.273	373.603	96.207	153.356	729	70.438	52.873
3	Macdonia and Thrace, Greece	Female	F: 13-17	15.676	1.094	337	757	108	649	14.582	13.079	0	0	1.302	201
4	Macdonia and Thrace, Greece	Female	F: 18-24	36.852	16.908	9.527	7.381	2.437	4.944	19.944	12.077	0	8	3.059	4.800
5	Eastern Macedonia and Thrace, Greece	Female	F: 25-34	52.528	18.002	9.864	8.138	2.545	5.593	34.526	25.156	0	8	4.361	5.001
6	Eastern Macedonia and Thrace, Greece	Female	F: 35-44	41.889	27.357	22.849	4.508	3918	590	14532	118	384	38	13394	598
7	Eastern Macedonia and Thrace, Greece	Female	F: 45-54	40.684	21.390	18.490	2900	2573	327	19.194	20	3166	61	15308	639
8	Eastern Macedonia and Thrace, Greece	Female	F: 55-64	36.500	8.275	7.458	817	802	15	28.225	0	11981	75	15477	692
9	Eastern Macedonia and Thrace, Greece	Female	F: 65+	73.282	483	476	7	7	0	72.799	0	62420	74	9205	1100
10	Eastern Macedonia and Thrace, Greece	Male	M: 13-17	17.630	2.065	701	1.364	205	1.159	15.565	13.740	0	0	0	1.825
11	Eastern Macedonia and Thrace, Greece	Male	M: 18-24	20.101	9.922	5.936	3.986	1.461	2.525	10.179	5.624	0	3	34	4.518
12	Eastern Macedonia and Thrace, Greece	Male	M: 25-34	41.894	37.431	29.305	8.126	5.316	2.810	4.463	1.327	0	42	65	3.029
13	Eastern Macedonia and Thrace, Greece	Male	M: 35-44	42.191	40.091	33.717	6374	5684	690	2100	102	612	93	60	1233

Εικόνα 4.15 Τα τροποποιημένα αρχεία για την οικογενειακή κατάσταση και τη κατάσταση ασχολίας των Ελλήνων.

4.3.2. Συσχέτιση Δεδομένων ΕΛΣΤΑΤ

Ο λόγος που χρησιμοποιούμε τα δεδομένα της ΕΛΣΤΑΤ είναι για να εξάγουμε μια επιπλέον δημογραφική πληροφορία για τους ακόλουθους της σελίδας μας πέρα από αυτές που μας παρέχει το Facebook. Αν η εφαρμογή αυτή είχε δημιουργηθεί 2 χρόνια πριν, όλα θα ήταν αρκετά πιο εύκολα από ότι τώρα. Με την θέσπιση του νόμου προστασίας δεδομένων αλλά και σε συνδυασμό με τα σκάνδαλα που εμπλέκεται το Facebook, σχετικά με την παροχή δεδομένων σε τρίτους, η παροχή δημόσιων πληροφοριών συρρικνώθηκε κατά πολύ. Παλιότερα μπορούσε κανείς να έχει πρόσβαση σε δημόσια δεδομένα-πληροφορίες ενός ατόμου (τα οποία ο ίδιος είχε θέσει ως δημόσια στο προφίλ του) και να γνωρίζει αν είναι π.χ. παντρεμένος, που έχει σπουδάσει, σε ποια πόλη μένει, που εργάζεται και άλλα. Πλέον η παροχή όλο αυτών των πληροφοριών έχει σταματήσει και δεν υπάρχει πια πρόσβαση σε τέτοια δεδομένα.

Έτσι ο μόνος τρόπος να εξάγουμε αυτές τις επιπλέον πληροφορίες που θέλουμε για τους ακόλουθους της σελίδας είναι να συσχετίσουμε τις πληροφορίες που έχουμε για αυτούς με τα δεδομένα της ΕΛΣΤΑΤ. Επειδή σχεδόν όλοι οι ακόλουθοι της σελίδας μένουν στην Ελλάδα και η

εφαρμογή αυτή απευθύνεται για ελληνικές σελίδες Facebook, το να συσχετίσουμε τα δεδομένα της σελίδας με τα δεδομένα της ΕΛΣΤΑΤ θα έχει νόημα και μπορεί να εξαχθεί ουσιαστική πληροφορία.

Αυτό που κάναμε για να συσχετίσουμε τα δεδομένα είναι κάτι αρκετά απλό και κατανοητό. Για κάθε κατηγορία-ομάδα που μας ενδιαφέρει, ανατρέχουμε στο αρχείο με τα κατάλληλα δεδομένα της ΕΛΣΤΑΤ και υπολογίζουμε το ποσοστό που της αναλογεί στο γεωγραφικό διαμέρισμα που έχουμε επιλέξει. Το ποσοστό αυτό το πολλαπλασιάζουμε με τον αριθμό των ακολουθών της σελίδας που υπάγονται στην κατηγορία-ομάδα που εξετάζουμε. Έτσι έχουμε μια αναλογική εκτίμηση για τα επίπεδα εκπαίδευσης, οικογενειακές καταστάσεις και ασχολίες των Ελλήνων ακόλουθων της σελίδας.

Όσο μεγαλύτερος ο αριθμός των ακόλουθων μιας σελίδας τόσο πιο κοντά σε κανονική κατανομή πληθυσμού είναι ο αριθμός των ακολούθων και το τόσο πιο κοντά στην πραγματικότητα είναι οι εκτιμήσεις για τους ακόλουθους και τα αποτελέσματα που παράγονται. Αλήθεια βέβαια είναι πως αν η σελίδα έχει μερικές δεκάδες ακόλουθους, ή απευθύνεται σε μία πολύ συγκεκριμένη και μικρή ομάδα ατόμων, τότε οι πληροφορίες που εξάγονται είναι δύσκολο να χρησιμοποιηθούν. Για παράδειγμα πάρα πολλές τιμές που έχουν αρκετή ποσοστιαία διαφορά (5% - 25%) όταν θα συσχετιστούν με 1 ή 2 ακόλουθους που μπορεί να έχει η σελίδα σε μια συγκεκριμένη κατηγορία, τότε η πληροφορία που θα παραχθεί δεν θα έχει κανένα νόημα. Και στις 2 περιπτώσεις το αποτέλεσμα θα είναι 0.

4.4. Γραφήματα Plotly – Dash

Η Plotly είναι μία εταιρία πληροφορικής που αναπτύσσει εργαλεία ανάλυσης και παρουσίασης δεδομένων. To Dash είναι ένα framework ανοιχτού κώδικα σχεδιασμένο για Python, που επιτρέπει τη δημιουργία web-based εφαρμογών ανάλυσης και παρουσίασης δεδομένων. Οι εφαρμογές που παράγονται μέσω του Dash μετατρέπονται σε Flask εφαρμογές ώστε εύκολα να μπορούν να τρέξουν σε online server και να χρησιμοποιηθούν από πολλά άτομα μέσω διαμοίρασης της URL διεύθυνσης της εφαρμογής. To Dash λειτουργεί σε όλες τις πλατφόρμες και είναι προσβάσιμο τόσο από desktop όσο και από κινητή συσκευή.

4.4.1. To Dashboard της Σελίδας

To Dashboard συνολικά απαρτίζεται από 11 γραφήματα τα οποία είναι χωρισμένα σε 4 κατηγορίες. Αυτές είναι οι εξής :

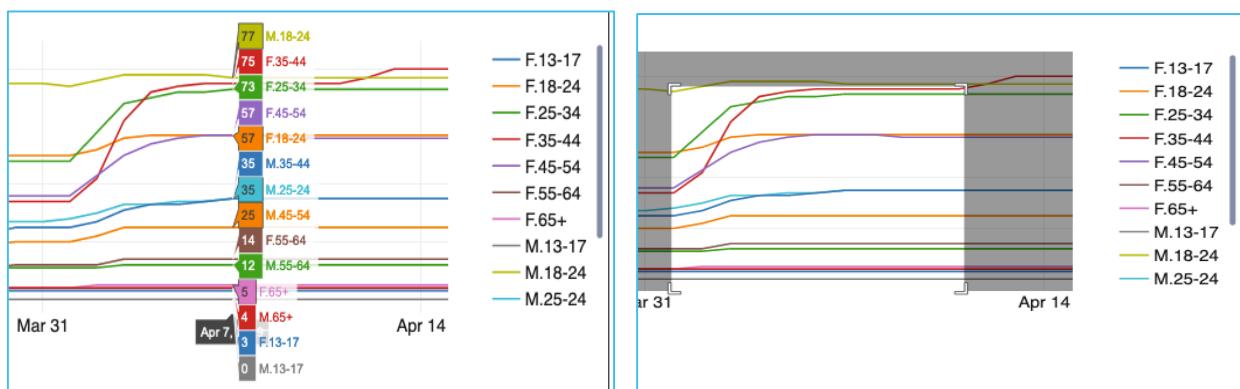
1. Πληροφορίες για τους ακόλουθους και στατιστικά της σελίδας.
2. Γεωγραφικές πληροφορίες και στατιστικά για τους ακόλουθους της σελίδας

3. Πληροφορίες και στατιστικά για κάθε ανάρτηση της σελίδας
4. Δημογραφική ανάλυση των αναρτήσεων για στρατηγική δαπανών.

Η πρώτη ομάδα γραφημάτων, “Information and stats about the fans of our page” όπως αναφέρεται και στην εφαρμογή, αποτελείται από 4 γραφήματα που αναλύουν τα δεδομένα για τους ακόλουθους της σελίδας και στατιστικά της σελίδας. Πιο συγκεκριμένα περιλαμβάνει τα εξής:

1. Ηλικία και φύλο για τους ακόλουθους της σελίδας

Εδώ προβάλλεται ένα line chart για τις ηλικίες των ακολούθων της σελίδας. Στον άξονα x έχουμε τις ημερομηνίες για τις οποίες έχουμε δεδομένα και στον άξονα y έχουμε τις τιμές για το πλήθος των ατόμων που ανήκουν σε κάθε ηλικιακή κατηγορία. Στο γράφημα υπάρχουν 14 κατηγορίες, 7 για κάθε φύλο, όπου κάθε κατηγορία αντιπροσωπεύει μια ηλικιακή ομάδα. Με μία κίνηση του ποντικιού πάνω από το γράφημα εμφανίζεται σε κάθε ημερομηνία, οι τιμές των ηλικιακών ομάδων και το πλήρες όνομα της ομάδας. Επειδή έχουμε συγκεντρώσει δεδομένα από αρκετές ημερομηνίες, ο χρήστης μπορεί να στοχεύσει ειδικότερα σε συγκεκριμένες μέρες, επιλέγοντας απλά το διάστημα που επιθυμεί. Αυτό γίνεται απλά τραβώντας ένα παραλληλόγραμμο με το ποντίκι στο επιθυμητό διάστημα. Μέσα από αυτό το διάγραμμά διακρίνουμε εύκολα την πορεία των ηλικιακών ομάδων των ακολούθων. Για συγκεκριμένες ημερομηνίες βλέπουμε σε ποιες ηλικιακές ομάδες έχει μεγαλύτερη απήχηση η σελίδα και προσέγγισε νέους ακόλουθους.

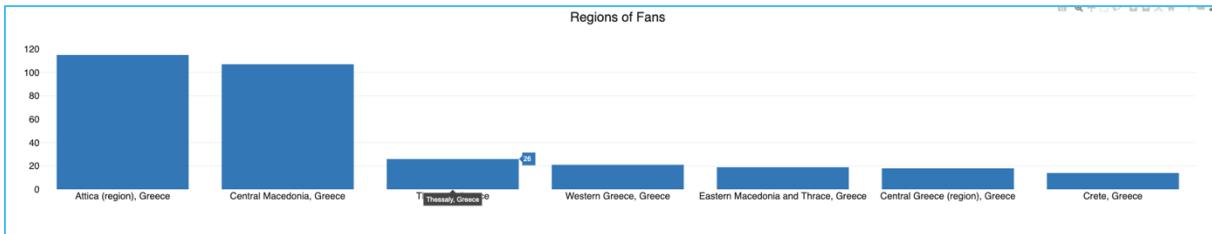


Εικόνα 4.16 Το γράφημα για τις ηλικίες των ακολούθων

2. Γεωγραφικά διαμερίσματα των ακολούθων.

Σε αυτό το γράφημα εμφανίζεται ένα bar chart με τα πρώτα 7 γεωγραφικά διαμερίσματα που μένουν οι ακόλουθοι. Οι τιμές που εμφανίζονται είναι οι πιο πρόσφατες τιμές που έχουμε συγκεντρώσει από το αρχείο των πόλεων και αντιστοιχούν στην τελευταία ημερομηνία που πάρθηκαν δεδομένα. Ο χρήστης έτσι έχει μία σαφή εικόνα για το πού μένουν γενικότερα οι ακόλουθοι του και σε ποιους απευθύνεται. Η εφαρμογή λειτουργεί εξίσου καλά και όταν πρόκειται για περιοχές του εξωτερικού. Στην προκείμενη περίπτωση

και οι 7 μπάρες αναπαριστούν μια ελληνική περιοχή, επειδή οι πλειοψηφία των ακολούθων μένει στην ελληνική επικράτεια.



Εικόνα 4.17 Το bar chart με τις περιοχές των ακολούθων.

3. Πληροφορίες για τη σελίδα.

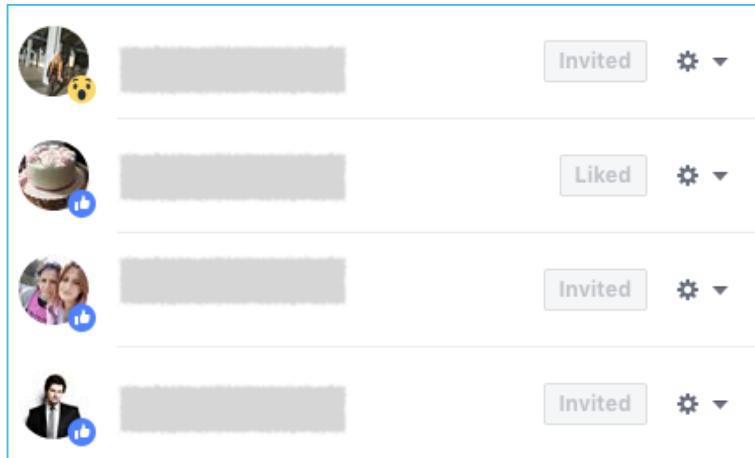
Μέσω αυτού του line chart γραφήματος προβάλουμε τις γενικές πληροφορίες που αφορούν την σελίδα μας. Στον άξονα x έχουμε τις ημερομηνίες για τις οποίες έχουμε δεδομένα και στον άξονα y έχουμε τις τιμές κάθε κατηγορίας δεδομένων. Αυτές είναι οι εξής:

- Προβολές της σελίδας (Page View Total)
- Ο αριθμός των ακόλουθων (Page Fans)
- Οι νέοι ακόλουθοι μέσω διαφημίσεων (Page Fans Adds Paid)
- Οι νέοι ακόλουθοι χωρίς διαφημίσεις (Page Fans Adds Non Paid)
- Οι εντυπώσεις της σελίδας σε κοινό [σε πόσα άτομα εμφανίστηκε η σελίδα στο timeline τους] (Page Impressions)
- Οι εντυπώσεις της σελίδας σε κοινό μέσω διαφημίσεων (Page Impressions Paid)
- Οι εντυπώσεις της σελίδας σε κοινό χωρίς διαφημίσεις (Page Impressions Non Paid)

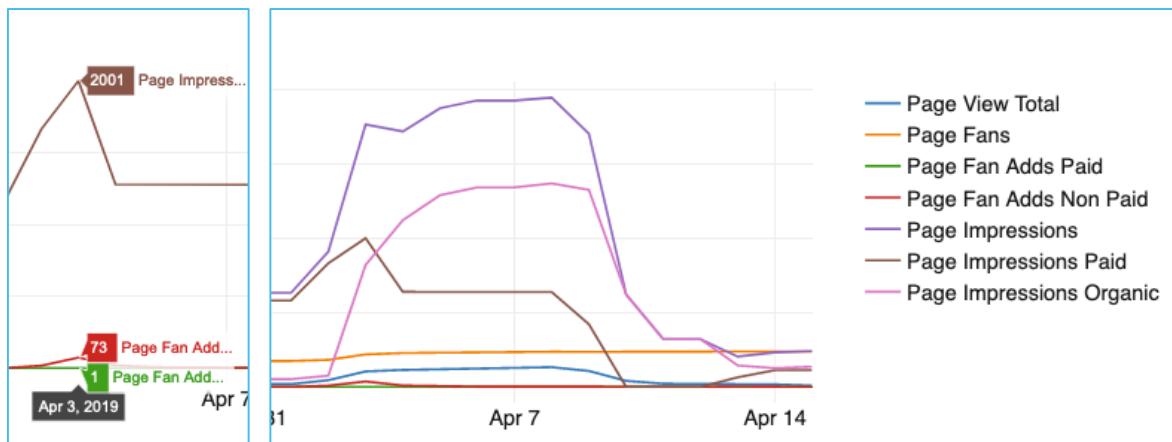
Όπως μπορεί να δει κανείς στο συγκεκριμένο γράφημα υπάρχει ένα παράδοξο. Στις ημερομηνίες που υπήρχε μεγάλη ακμή του Page Impressions Paid, δηλαδή δημοσιεύσαμε μια ανάρτηση μέσω διαφήμισης, οι νέες προσθήκες ακολούθων μέσω διαφήμισης στην σελίδα είναι μηδενικές. Χαρακτηριστικά στις 3/4/19 το Page Impressions Paid είναι 2000, Page Fans Adds Non Paid στο 73 και το Page Fans Adds Paid μόλις 1.

Θα έλεγε κανείς πως η διαφήμιση ήταν αποτυχία με μόνο 1 προσθήκη. Αυτή θα ήταν μία λάθος εκτίμηση καθώς οι 73 νέες προσθήκες προήλθαν και αυτές από την διαφήμιση, απλά με άλλο τρόπο. Όταν ένα άτομο (που δεν είναι ακόλουθος) πατήσει «Μου αρέσει – like» ή κάνει οποιαδήποτε άλλη αντίδραση στην διαφημιζόμενη ανάρτηση, τότε δίνεται στον διαχειριστή της σελίδας η δυνατότητα να προσκαλέσει (invite) το άτομο αυτό να πατήσει «Μου αρέσει – like» στη σελίδα και να γίνει ακόλουθος. Αν αποδεχτεί το άτομο το αίτημα,

τότε η νέα προσθήκη προσμετράται στις Page Fans Adds Non Paid και όχι στις Page Fans Adds Paid, που λογικά θα έπρεπε. Στις περισσότερες περιπτώσεις που διαφημίζουμε μία ανάρτηση συμβαίνει αυτό το φαινόμενο, οπότε θα πρέπει να είμαστε προσεχτικοί στις τελικές αξιολογήσεις και στα συμπεράσματα.



Εικόνα 4.18 Η δυνατότητα να προσκαλέσει ο διαχειριστής νέα άτομα στη σελίδα.



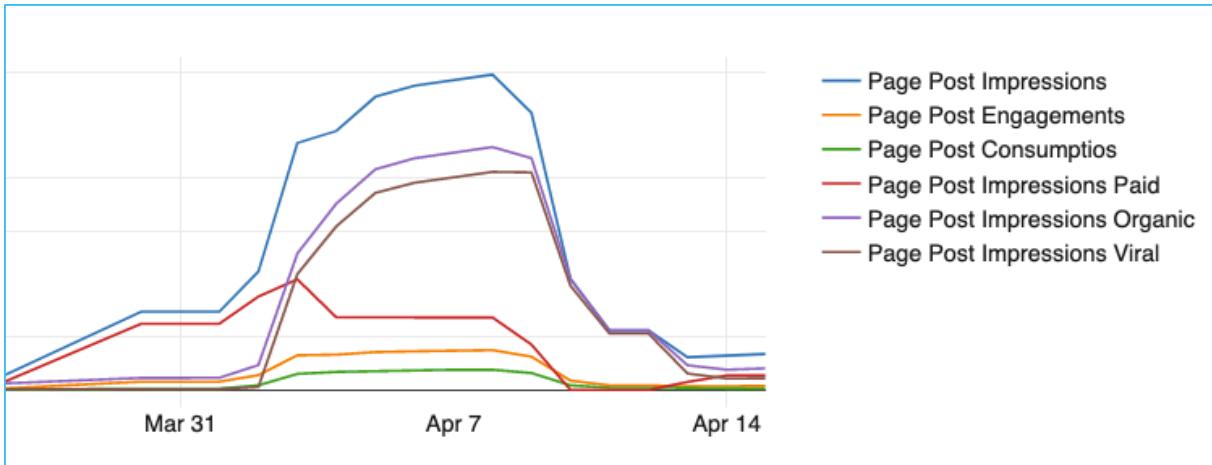
Εικόνα 4.19 Το line chart με τις πληροφορίες για την σελίδα

4. Πληροφορίες για τις αναρτήσεις της σελίδας

Εδώ παρουσιάζεται ένα line chart που αφορά τις πληροφορίες και τα δεδομένα που έχουμε σχετικά με τις αναρτήσεις της σελίδας. Όπως και στα προηγούμενα στο άξονα x βρίσκονται οι ημερομηνίες για τις οποίες έχουμε δεδομένα και στο άξονα y βρίσκονται οι τιμές κάθε κατηγορίας δεδομένων. Αυτές είναι οι ακόλουθες :

- Εντυπώσεις των αναρτήσεων της σελίδας [σε πόσα άτομα εμφανίστηκε κάποια ανάρτηση της σελίδα στο timeline τους] (Page Post Impressions)
- Διαδράσεις ατόμων στις αναρτήσεις της σελίδας (Page Post Engagements)
- Κλικς που είχε οποιαδήποτε ανάρτηση της σελίδας (Page Post Consumptions)

- Εντυπώσεις των αναρτήσεων της σελίδας μέσω διαφημίσεων (Page Post Impressions Paid)
- Εντυπώσεις των αναρτήσεων της σελίδας από τους ακόλουθους (Page Post Impressions Organic)
- Εντυπώσεις των αναρτήσεων της σελίδας από φίλους των ακολούθων (Page Post Impressions Viral)



Εικόνα 4.20 Το line chart με τις πληροφορίες για τις αναρτήσεις της σελίδας

Αυτά τα 4 γραφήματα αποτελούν την πρώτη ομάδα που αφορά τις γενικές πληροφορίες για την σελίδα. Στη συνέχεια ακολουθούν οι γεωγραφικές πληροφορίες που αφορούν τους ακόλουθους της σελίδας. Αυτές οι πληροφορίες αναπαρίστανται στα εξής 3 γραφήματα :

1. Γράφημα πόλεων

Σε αυτό το γράφημα ο χρήστης μπορεί να δει τον αριθμό των ακόλουθων που μένουν σε μια συγκεκριμένη πόλη στο διάστημα των ημερομηνιών που υπάρχουν δεδομένα. Παράλληλα ο χρήστης μπορεί να συγκρίνει τα δεδομένα αρκετών πόλεων, καθώς υπάρχει η δυνατότητα στο dropdown menu να επιλέξει παραπάνω από μία πόλη. Έτσι η σύγκριση των δεδομένων μεταξύ πόλεων απλοποιείται αρκετά. Οι πόλεις που εμφανίζονται στο μενού αποτελούν όλες τις πόλεις που υπάρχουν στο excel αρχείο. Αυτό σημαίνει πως έχουμε όλες τις πόλεις από τις οποίες για οποιαδήποτε χρονική στιγμή είχαμε έναν ακόλουθο που έμενε εκεί και όχι μόνο αυτές στις οποίες μένουν ακόλουθοι την παρούσα χρονική στιγμή.

Ο χρήστης αρκεί να επιλέξει ορισμένες πόλεις που θέλει να συγκρίνει από το μενού ή να τις πληκτρολογήσει με αγγλικούς χαρακτήρες και να πατήσει submit για να ανανεωθεί το γράφημα. Πληκτρολογώντας το όνομα ενός γεωγραφικού διαμερίσματος, εμφανίζεται στο μενού μόνοι οι πόλεις που ανήκουν σε αυτό το διαμέρισμα. Κάνοντας κλικ δεξιά στο όνομα κάποιας πόλης, αυτόματα εμφανίζεται στο γράφημα μόνο η συγκεκριμένη πόλη. Κάνοντας

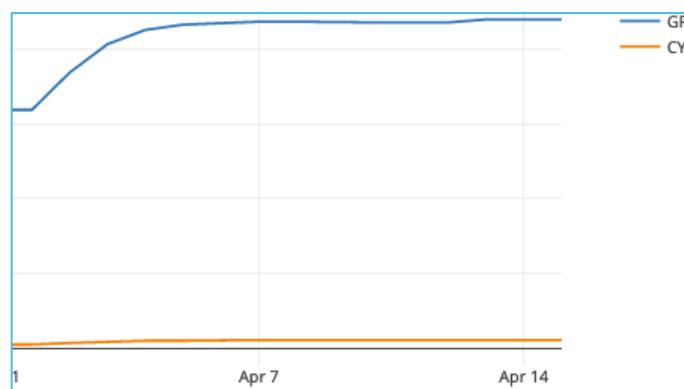
κλικ και σε άλλες τότε εμφανίζονται μόνο οι συγκεκριμένες. Οι τιμές και οι κλίμακες του άξονα γ αλλάζουν μαζί με τις επιλογές που γίνονται, για να παρουσιαστούν σωστά οι αναλογίες μεταξύ των πόλεων.



Εικόνα 4.21 Το διαδραστικό line chart πόλεων

2. Γράφημα χωρών

Εδώ εμφανίζουμε τον αριθμό των ακολούθων που μένουν σε κάθε χώρα στο διάστημα των ημερομηνιών που υπάρχουν δεδομένα. Όπως και στο γράφημα πόλεων, έτσι και εδώ ο χρήστης μπορεί να επιλέξει μία ή παραπάνω χώρες να εμφανιστούν στο γράφημα και να συγκρίνει τις τιμές τους. Οι χώρες στο μενού εμφανίζονται με τον διεθνή κωδικό τους.

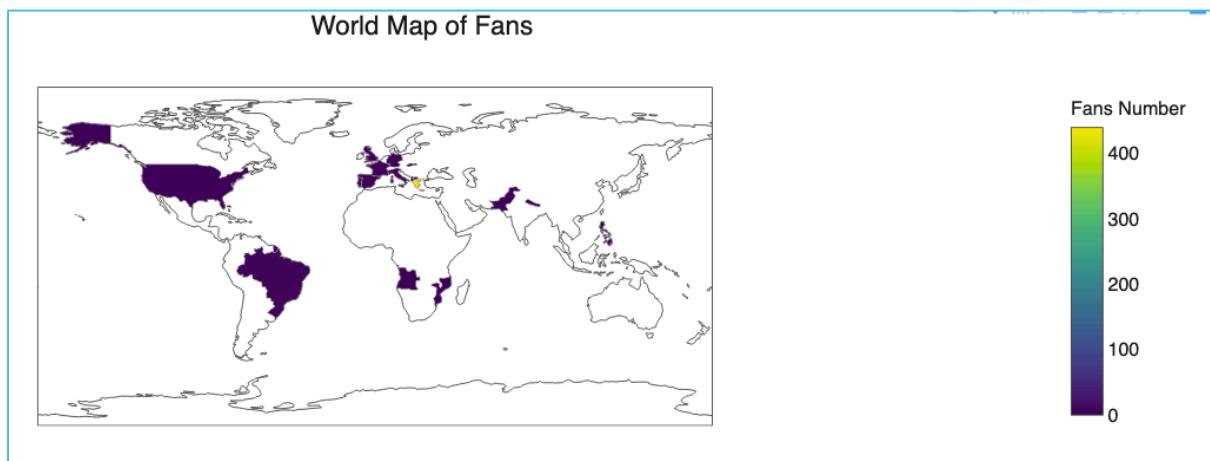


Εικόνα 4.22 Το διαδραστικό line chart χωρών

3. Παγκόσμιος χάρτης ακολούθων

Σε αυτό το γράφημα εμφανίζουμε έναν θερμό παγκόσμιο χάρτη (heated world map). Σε αυτό τον χάρτη οι χώρες στις οποίες μένουν ακόλουθοι εμφανίζονται με ένα ορισμένο χρώμα. Το

χρώμα αυτό εξαρτάται από το πλήθος των ακολούθων που έχει κάθε χώρα και αλλάζει σταδιακά σε gradient αποχρώσεις. Εμείς έχουμε ορίσει για τις χαμηλές τιμές σκούρες αποχρώσεις του μπλε, ενώ για τις μεγάλες τιμές ορίσαμε ανοιχτές αποχρώσεις του κίτρινου. Οι αποχρώσεις προσαρμόζονται αυτόματα στο εύρος των τιμών. Στο δεξί μέρος του γραφήματος υπάρχει χρωματικό υπόδειγμα που πληροφορεί τον χρήστη για τις τιμές που αντιστοιχεί κάθε χρώμα. Βέβαια και με ένα από πέρασμα του ποντικιού πάνω από κάθε χώρα στο χάρτη εμφανίζεται ο ακριβής αριθμός των ακόλουθων.



Εικόνα 4.23 Ο παγκόσμιος χάρτης ακολούθων

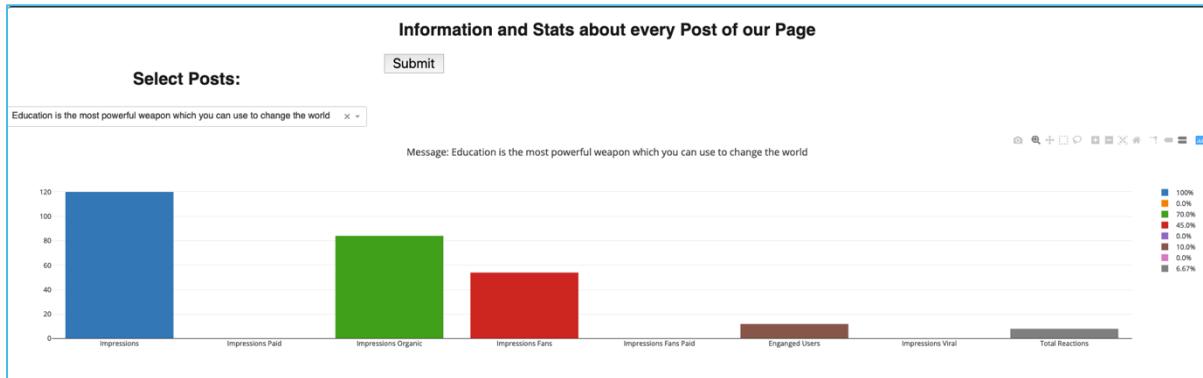
Αυτά τα 3 γραφήματα αποτελούν την δεύτερη ομάδα που αφορά τις γεωγραφικές πληροφορίες για τους ακόλουθους της σελίδα. Η επόμενη ομάδα περιέχει μόνο ένα γράφημα, το οποίο παρουσιάζει πληροφορίες για τις αναρτήσεις της σελίδας. Αυτό είναι το εξής :

1. Πληροφορίες και στατιστικές για κάθε ξεχωριστή ανάρτηση της σελίδας

Μέσω αυτού του διαδραστικού bar chart παρουσιάζονται τα δεδομένα που έχουμε συγκεντρώσει για την κάθε ανάρτηση ξεχωριστά από την ημερομηνία που δημιουργήθηκε η συγκεκριμένη ανάρτηση μέχρι και σήμερα. Ο χρήστης επιλέγει από το dropdown menu την ανάρτηση που τον ενδιαφέρει και πατάει submit για να ενημερωθεί το γράφημα με τα σωστά δεδομένα. Τα δεδομένα που παρουσιάζουμε είναι τα εξής:

- Εντυπώσεις (Impressions)
- Εντυπώσεις μέσω διαφήμισης (Impressions Paid)
- Εντυπώσεις χωρίς διαφήμιση (Impressions Organic)
- Εντυπώσεις ανάμεσα στους ακόλουθους (Impressions Fans)
- Εντυπώσεις ανάμεσα στους ακόλουθους μέσω διαφήμισης (Impressions Fans Paid)
- Άτομα που έχουν κάποια διάδραση με την συγκεκριμένη ανάρτηση (Engaged Users)

- Εντυπώσεις σε φίλους ακολούθων (Impressions Viral)
- Συνολικές αντιδράσεις (Total Reactions)



Εικόνα 4.24 Το γράφημα με τις πληροφορίες για κάθε ανάρτηση

Η επόμενη ομάδα γραφημάτων που ακολουθεί είναι ουσιαστικά το σημαντικότερο κομμάτι του dashboard και της εργασίας. Σε αυτό το κομμάτι γίνεται δημογραφική ανάλυση των δεδομένων μας για να καθορίσουμε μια στρατηγική επένδυσης σε διαφημιζόμενες αναρτήσεις. Σκοπός μας είναι να καθορίσουμε το target group ατόμων ή ακολούθων, στο οποίο η ανάρτηση μας θα έχει τη μέγιστη απόδοση ή να ανακαλύψουμε σε ποια ομάδα – target group απευθύνεται κάθε ανάρτηση συγκεκριμένα. Παράλληλα μπορούμε να χρησιμοποιήσουμε το εργαλείο αυτό για να βελτιώσουμε την διαφημιστική καμπάνια μιας σελίδας Facebook αποτρέποντας την από το να ξοδέψει χρήματα σε ομάδες ατόμων που πρακτικά δεν χρειάζεται και δεν θα την αποφέρει νέους ακόλουθους και εν δυνάμῃ πελάτες.

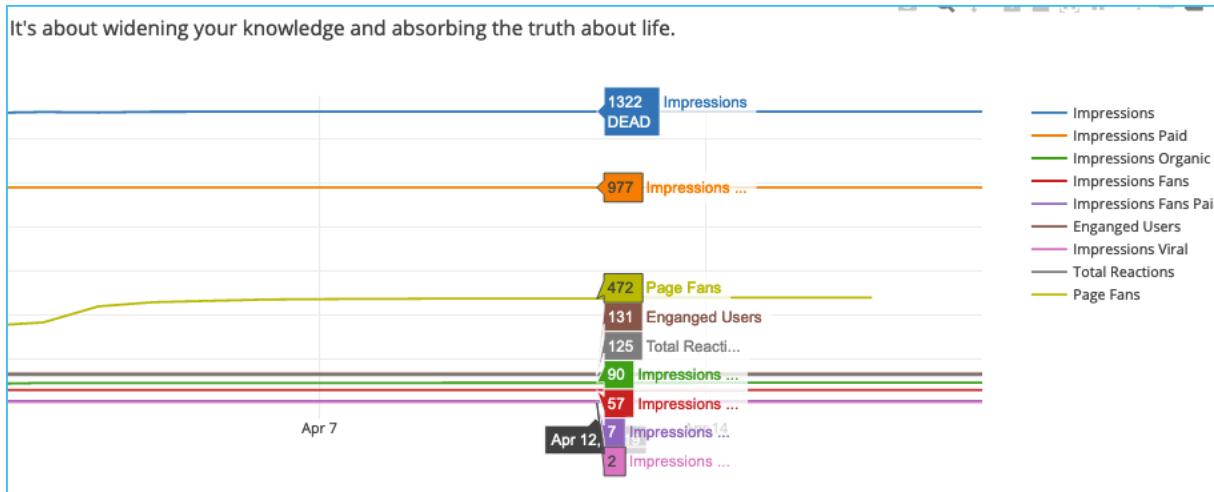
Η δημογραφική ανάλυση των αναρτήσεων αποτελείται από 5 στοιχεία 3 γραφήματα και 2 διαδραστικά κείμενα, όπου το ένα συμπληρώνει το άλλο για την εξαγωγή πληροφορίας.

1. Lines of Each Post:

Σε αυτό το γράφημα παρουσιάζουμε όπως και ακριβώς πριν τις πληροφορίες για κάθε ανάρτηση ξεχωριστά με την μόνη διαφορά πως εμφανίζεται επιπλέον και ο αριθμός των ακολούθων της σελίδας. Αυτό κάνει ευκολότερο τον εντοπισμό συσχετίσεων μεταξύ αυξήσεων στον αριθμό των εντυπώσεων (Impressions) μιας ανάρτησης και του συνολικού αριθμού ακολούθων. Επιπλέον κατά την κίνηση του ποντικιού πάνω από το γράφημα εμφανίζεται στην γραμμή των εντυπώσεων 3 ενδείξεις: START, ALIVE, DEAD.

- **START:** Συμβολίζουμε την ημέρα δημιουργίας της ανάρτησης και θέτουμε ως αφετηρία των εντυπώσεων τον αριθμό που συγκέντρωσε εκείνη τη μέρα.
- **ALIVE:** Συμβολίζουμε όλες τις επόμενες μέρες που η πορεία των εντυπώσεων της ανάρτησης είναι ανοδική και ο αριθμός τους αυξάνεται.

- DEAD: Συμβολίζουμε τις ημερομηνίες όπου ο αριθμός των εντυπώσεων της ανάρτησης έμεινε στάσιμος. Δηλαδή κανένα άτομο δεν είδε την συγκεκριμένη ανάρτηση εκείνη την ημέρα.



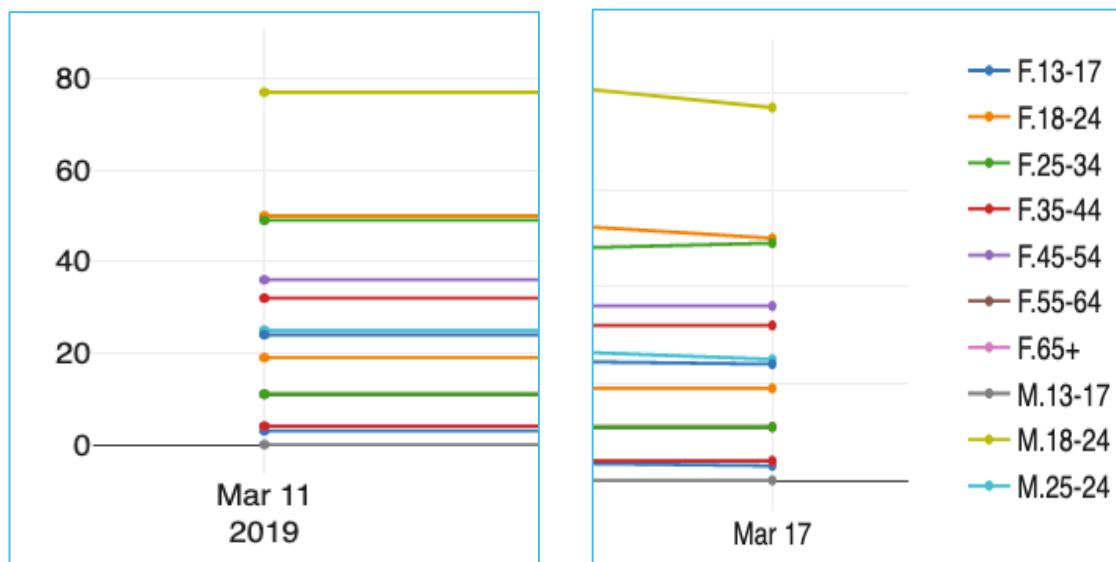
Εικόνα 4.25 Οι στατιστικές κάθε ανάρτησης

2. Στη συνέχεια εμφανίζεται σε ένα πλαίσιο κειμένου οι ημερομηνίες στις οποίες η ανάρτηση έχει αυξανόμενο αριθμό εντυπώσεων (Impressions), δηλαδή κάθε μέρα περισσότερα άτομα σε σχέση με χθες βλέπουν την ανάρτηση. Πιο συγκεκριμένα εμφανίζονται 2 ημερομηνίες: η ημέρα που η ανάρτηση δημιουργήθηκε, δηλαδή η αφετηρία (START) και η πρώτη ημέρα όπου η ανάρτηση σταμάτησε να έχει αύξων αριθμό εντυπώσεων (DEAD). Αυτό το διάστημα είναι η χρονική περίοδος, όπου η ανάρτηση είναι «ζωντανή» και εξακολουθεί να συγκεντρώνει εντυπώσεις. Πολλές φορές μία ανάρτηση μπορεί να σταματήσει να έχει νέες εντυπώσεις και μετά από λίγες μέρες να συνεχίσει να αυξάνεται ο αριθμός των εντυπώσεων της. Αυτό συμβαίνει εξαιτίας κάποιας νέας ανάρτησης που φέρνει νέους επισκέπτες στη σελίδα και ως αποτέλεσμα αυτοί βλέπουν και τις προηγούμενες αναρτήσεις που υπάρχουν. Έτσι αν λαμβάναμε υπόψιν κάποια άλλη ημερομηνία πέρα της πρώτης όπου δεν αυξάνονται οι εντυπώσεις, θα είχαμε λάθος εκτίμηση για την ζωή της ανάρτησης και την δυναμικότητα που είχε. Σε περίπτωση που η ανάρτηση είναι αρκετά πρόσφατη και δεν υπάρχει ημερομηνία στάσιμη πορεία εντυπώσεων και "DEAD" ένδειξη, τότε ως τελική ημέρα θεωρούμε την τελευταία ημερομηνία που έχουμε δεδομένα. Τέλος οι ημερομηνίες εμφανίζονται σε μορφή ISO, όπου στην αρχή εμφανίζονται έτος-μήνας-ημέρα, ακολουθεί το Τ και στη συνέχεια ώρα-λεπτά-δευτερόλεπτα.

Post is alive from: 2019-03-11T00:00:00.000000000 until: 2019-03-17T00:00:00.000000000

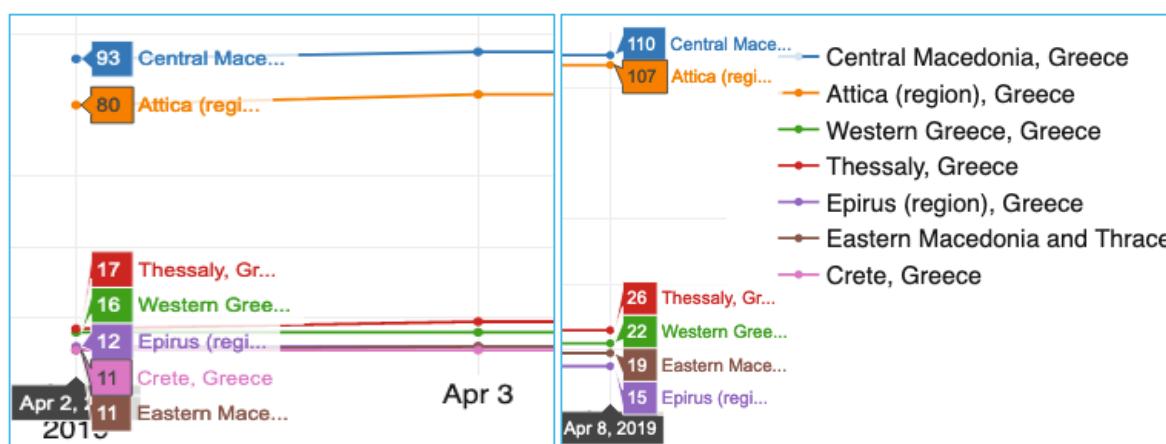
Εικόνα 4.26 Το διάστημα όπου η συγκεκριμένη ανάρτηση είναι ζωντανή (alive)

3. Έπειτα έχοντας τις συγκεκριμένες ημερομηνίες για τις οποίες η ανάρτηση ήταν ενεργή, μπορούμε να δούμε ακριβώς τι τιμές είχαν τα δημογραφικά δεδομένα της σελίδας και να εξετάσουμε σε ποια ηλικιακή ομάδα υπήρξε η μεγαλύτερη απήχηση και ως αποτέλεσμα αύξηση στον αριθμό των ακολούθων. Το γράφημα που προβάλλεται αποτελεί ένα υπόγραφημα του ηλικία-φύλου γραφήματος που υπάρχει παραπάνω, στο οποίο το εύρος ημερομηνιών αλλάζει δυναμικά κάθε φορά που επιλέγεται άλλη ανάρτηση.



Εικόνα 4.27 Το γράφημα ηλικιών-φύλου στις συγκεκριμένες ημερομηνίες κάθε ανάρτησης.

4. Παρομοίως λειτουργεί και το επόμενο γράφημα που αφορά τα γεωγραφικά διαμερίσματα των ακολούθων. Εδώ παρουσιάζονται όπως και παραπάνω αλλά σε line chart αντί bar chart, τα 7 πρώτα σε πλήθος ακολούθων γεωγραφικά διαμερίσματα. Η διαφορά είναι πως το εύρος ημερομηνιών που προβάλλεται είναι αυτό που εμφανίζεται παραπάνω στο διάστημα που η ανάρτηση είναι ζωντανή. Αυτό μας επιτρέπει να εξετάσουμε σε ποια γεωγραφικά διαμερίσματα είχε η συγκεκριμένη ανάρτηση την μεγαλύτερη απήχηση και απέφερε τους περισσότερους νέους ακόλουθους.



Εικόνα 4.28 Το line chart για τα γεωγραφικά διαμερίσματα σε συγκεκριμένες ημερομηνίες

5. Στο τελευταίο μέρος της ανάλυσης για στρατηγική δαπανών υπάρχουν 2 πλαίσια κειμένου στα οποία παρουσιάζεται η ηλικιακή ομάδα και το γεωγραφικό διαμέρισμα, στο οποίο η συγκεκριμένη ανάρτηση είχε τα καλύτερα αποτελέσματα. Ο αλγόριθμος ελέγχει όλες τις διαθέσιμες τιμές για κάθε ηλικιακή ομάδα και γεωγραφικό διαμέρισμα και βρίσκει τη μέγιστη διαφορά μεταξύ των τιμών. Αφού εντοπίσει την ηλικιακή ομάδα και το γεωγραφικό διαμέρισμα με τη μέγιστη διαφορά, εμφανίζει το όνομα της συγκεκριμένης ηλικιακής ομάδας και γεωγραφικού διαμερίσματα, την διαφορά στις τιμές από την αφετηρία μέχρι το τέλος και την ποσοστιαία αύξηση. Έτσι ο χρήστης εύκολα γνωρίζει το target group με τη βέλτιστη απήχηση σε κάθε ανάρτηση ξεχωριστά.

Biggest Growth at Ages: F.35-44 from 42 to 75 [78.57%]

Biggest Growth at Regions: Attica (region), Greece from 80 to 107 [33.75%]

Εικόνα 4.29 To target group με τη μεγαλύτερη απήχηση για κάθε ανάρτηση

Συνοψίζοντας με τη χρήση των παραπάνω γραφημάτων που προβάλουν πληροφορίες για τη σελίδα, τις αναρτήσεις και τους ακόλουθους, ο χρήστης μπορεί να αντλήσει χρήσιμες πληροφορίες και γνώση για την σελίδα του, και να ανακαλύψει τα βέλτιστα target group ατόμων για τις διαφημιστικές καμπάνιες.

Σύνοψη Κεφαλαίου

Στο κεφάλαιο αυτό παρουσιάσαμε το πρακτικό κομμάτι της εργασίας, που αποτελείται από την ανάπτυξη κώδικα, την διαχείριση αρχείων και την εξαγωγή αποτελεσμάτων. Για την ανάπτυξη του κώδικα χρησιμοποιήθηκε η γλώσσα Python, η οποία εύκολα μας επιτρέπει να ζητάμε δεδομένα για την σελίδας μας από το Graph API του Facebook.

Τα δεδομένα που παίρνουμε για την σελίδα μας, τους ακόλουθους μας και τα δημογραφικά τους στοιχεία τα αποθηκεύουμε σε αρχεία excel. Έπειτα αυτά τα αρχεία χρησιμοποιούνται από την εφαρμογή για να εξαγάγουμε πληροφορίες και στατιστικά δεδομένα. Με αυτό τον τρόπο έχουμε μια λεπτομερής ανάλυση των δεδομένων της σελίδας μας και έχουμε πλήρη επίγνωση σε τι κοινό απευθυνόμαστε αλλά και πόση απήχηση έχουμε.

Για τα παραπάνω δεδομένα υπάρχει η δυνατότητα συσχέτισης με τα δημογραφικά δεδομένα της Ελληνικής Στατιστικής Υπηρεσίας (ΕΛΣΤΑΤ), ώστε να προσφέρουμε μεγαλύτερη επίγνωση στον διαχειριστή της σελίδας σχετικά με τους ακόλουθους. Γνωρίζοντας το μορφωτικό, το κοινωνικό και το επίπεδο ασχολίας των ανθρώπων μιας περιοχής, είναι εύκολο μετά να ρυθμίσει κανείς αναλόγως την διαφημιστική του καμπάνια για στοχευμένα και καλύτερα αποτελέσματα.

Αυτή τη επιπλέον γνώση προσφέρει και η εφαρμογή μας. Διαβάζοντας καλύτερα το κοινό στο οποίο απευθυνόμαστε, μπορούμε να κρίνουμε και να επιλέξουμε την ομάδα ατόμων που πρέπει να εστιάσουμε ώστε να έχουμε τα βέλτιστα αποτελέσματα. Με άλλα λόγια από το πανελλαδικό επίπεδο συγκλίνουμε σε ένα στενά ορισμένο target group, το οποίο θα μας επιφέρει τους περισσότερους νέους ακόλουθους και κατά επέκταση τους περισσότερους νέους πελάτες.

Η ομάδα ατόμων στην οποία θα εστιάσουμε την διαφημιστική καμπάνια χαρακτηρίζεται από ένα συγκεκριμένο φύλο, ενός συγκεκριμένου εύρους ηλικίας που διαμένει σε μία συγκεκριμένη γεωγραφική περιοχή.

Κεφάλαιο 5

Συμπεράσματα

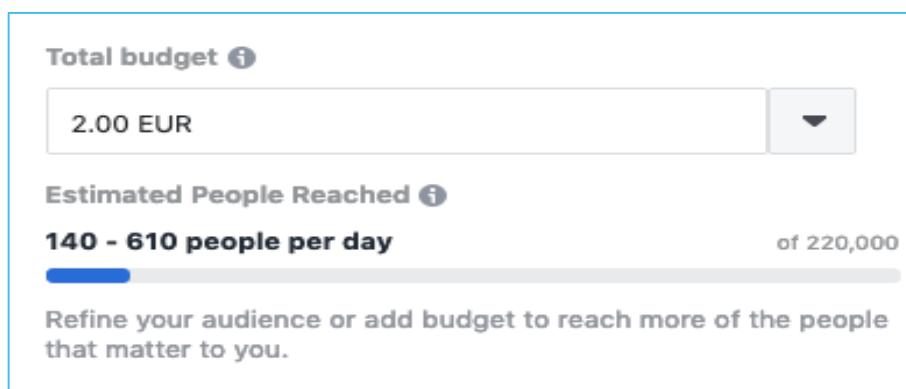
Για να ελέγξουμε κατά πόσο η εφαρμογή που αναπτύξαμε είχε νόημα, πετυχαίνει τους στόχους της και μπορεί να χρησιμοποιηθεί από οποιονδήποτε χρήστη-διαχειριστή σελίδας Facebook, κάναμε το ακόλουθο πείραμα στη demo σελίδα, My Thesis Page που φτιάχαμε.

Σκοπός του πειράματος είναι να εξετάσουμε την απήχηση μίας διαφημιζόμενης ανάρτησης και στη συνέχεια κάνοντας ανάλυση των αποτελεσμάτων να βελτιστοποιήσουμε την διαφημιστική καμπάνια μας. Στην ουσία θα ανεβάσουμε μια δημοσίευση στη σελίδα μας, την οποία αρχικά θα την διαφημίζαμε πανελλαδικά και σε όλες τις ηλικιακές ομάδες, θα αναλύσουμε τα δεδομένα της ανάρτησης, θα επιλέξουμε το νέο-βέλτιστο target group για την ανάρτηση και σε τελική φάση θα επενδύσουμε μέσω διαφήμισης στο νέο target group.

Ουσιαστικά ένα εργαλείο σαν το δικό μας, βοηθάει τις σελίδες να μην σπαταλούν χρήματα για διαφημίσεις σε άτομα, στα οποία η σελίδα δεν έχει απήχηση και που δεν τους συμφέρει.

Αρχικά θέτουμε το budget που έχουμε για να επενδύσουμε στην διαφημιζόμενη ανάρτηση. Η τακτική είναι σε πρώτη φάση να δαπανηθούν λίγα χρήματα σε μία αναγνωριστική ανάρτηση, να γίνει η ανάλυση των αποτελεσμάτων και τέλος να χρησιμοποιηθεί όλο το υπόλοιπο budget στο target group που θα μας επιφέρει το καλύτερο αποτέλεσμα και τους περισσότερους νέους πελάτες.

Το budget θα μοιραστεί ως εξής: Στην πρώτη αναγνωριστική διαφήμιση θα χρησιμοποιήσουμε 2 € για 2 μέρες. Δηλαδή θα ξοδεύουμε 1 €/ημέρα. Αυτό σύμφωνα με το Facebook σημαίνει πως απευθυνόμαστε σε 140 έως 610 άτομα ημερησίως πανελλαδικά. Συγκριτικά για το μέγεθος της σελίδας, το 140-610 είναι ένα αρκετά καλό νούμερο, για να πάρουμε ενδεικτικά πρώτα αποτελέσματα. Με το ίδιο σταθερό ποσό μπορεί κανείς να ρυθμίσει τον τρόπο διαφήμισης και να προσεγγίσει περισσότερα άτομα σε λιγότερες μέρες ή λιγότερα άτομα σε περισσότερες μέρες.

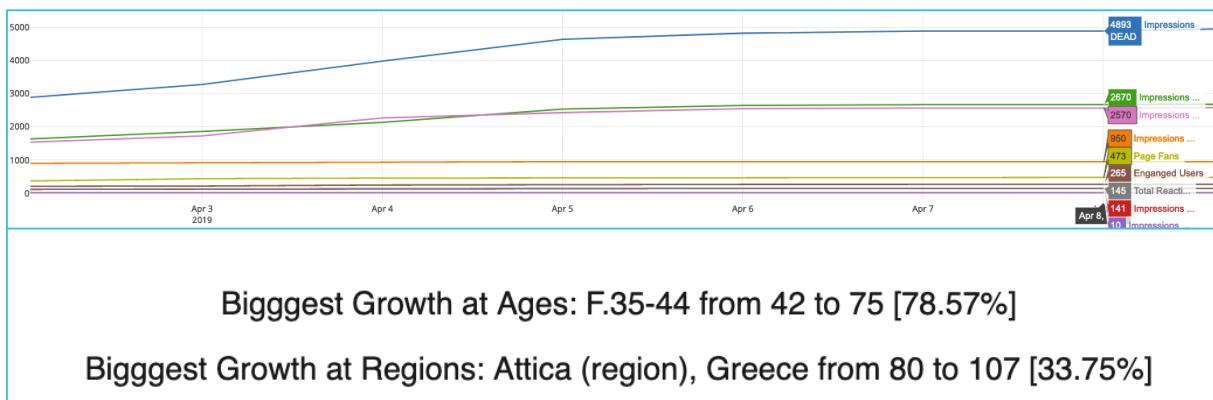


Εικόνα 5.1 Ορίζουμε το budget της διαφημιζόμενης ανάρτησης.

Αφού ορίσουμε το ποσό που θα ξοδέψουμε στη διαφήμιση και το target group μας, περιμένουμε από το Facebook να εξετάσει και να επικυρώσει την ανάρτηση μας ώστε να αρχίσει να κυκλοφορεί στον κόσμο.

Μετά από δύο μέρες (2-4-2019 έως 4-4-2019) που διαφημίζαμε την ανάρτηση πήραμε τα εξής αποτελέσματα :

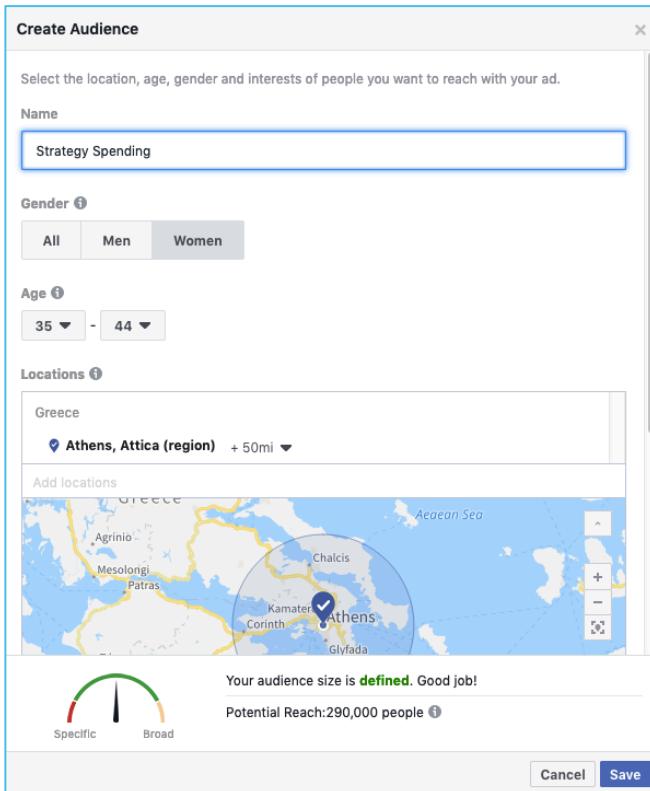
- Η ανάρτηση ήταν ζωντανή από 2-4-2019 έως 8-4-2019 (4893 εντυπώσεις)
- Συνολικά 186 αντιδράσεις
- Στο διάστημα που διαφημίζόταν είχε αρχικά 2873 εντυπώσεις και την τελευταία μέρα 3987 εντυπώσεις
- Οι ακόλουθοι της σελίδας αυξήθηκαν από 364 σε 444 άτομα
- Η ηλικιακή ομάδα με την μεγαλύτερη αύξηση ήταν οι γυναίκες 35-44 (F.35-44), από 42 σε 75 άτομα (33 νέοι ακόλουθοι 78,57%)
- Η γεωγραφική περιοχή με την μεγαλύτερη αύξηση ήταν η Αττική, από 80 σε 107 άτομα (27 νέοι ακόλουθοι 33.75%)



Εικόνα 5.2 Τα αποτελέσματα της διαφημιζόμενης ανάρτησης.

Με αυτές τις πληροφορίες, ξέρουμε τώρα ποιο είναι το βέλτιστο target group για την συγκεκριμένη ανάρτηση και που ακριβώς χρειάζεται να επενδύσουμε τα υπόλοιπα χρήματα του budget μας. Έτσι μετά από λίγες μέρες αναρτήσαμε ξανά την ίδια δημοσίευση, με στοχευμένη διαφήμιση στο target group που θέλουμε. Περιμέναμε σχεδόν 10 ημέρες από την πρώτη ανάρτηση για να ανεβάσουμε την καινούργια, καθώς υπήρχαν μοιράσματα και σχόλια στην ανάρτηση και συνεχώς νέοι ακόλουθοι -εντυπώσεις, πράγμα που θα επηρέαζε την ακρίβεια των νέων μετρήσεων.

Η νέα ανάρτηση διαφημίστηκε στοχευμένα σε γυναίκες ηλικίας 35 έως 44 ετών που μένουν στην ευρύτερη περιοχή της αττικής.



Εικόνα 5.3 Η νέα διαφημιζόμενη ανάρτηση στο συγκεκριμένο target group που ορίσαμε.

Τα αποτελέσματα που έχουμε από την δεύτερη διαφημιζόμενη ανάρτηση είναι τα εξής :

- Η ανάρτηση ήταν ζωντανή από 20-4-2019 έως 2-5-2019 (6302 εντυπώσεις)
- Συνολικά 224 αντιδράσεις
- Στο διάστημα που διαφημιζόταν είχε αρχικά 2268 εντυπώσεις και την τελευταία μέρα 4405 εντυπώσεις
- Οι ακόλουθοι της σελίδας αυξήθηκαν από 478 σε 584 άτομα
- Η ηλικιακή ομάδα με την μεγαλύτερη αύξηση ήταν οι γυναίκες 35-44 (F.35-44), από 79 σε 129 άτομα (50 νέοι ακόλουθοι 63.29 %)
- Η γεωγραφική περιοχή με την μεγαλύτερη αύξηση ήταν η Αττική, από 124 σε 163 άτομα (39 νέοι ακόλουθοι 31.45%)

Biggest Growth at Ages: F.35-44 from 79 to 129 [63.29%]

Biggest Growth at Regions: Attica (region), Greece from 124 to 163 [31.45%]

Εικόνα 5.3 Τα αποτελέσματα της νέας διαφημιζόμενης ανάρτησης

Βλέπουμε πως σε μικρότερο διάστημα ημερών τα αποτελέσματα είναι καλύτερα. Οι ακόλουθοι στις γυναίκες 35-44 αυξήθηκαν κατά 50 άτομα και στην περιοχή της αττικής κατά 39 άτομα.

Συνολικά στις 3 μέρες που διαφημίζοταν η ανάρτηση είχαμε 106 νέους ακόλουθους στην σελίδα μας. Συγκριτικά με την πρώτη διαφημίζόμενη ανάρτηση έχουμε 26 περισσότερους νέους ακόλουθους. Για να εκτιμηθούν πλήρως τα αποτελέσματα για την απήχηση μίας ανάρτησης και την επιτυχία που έχει πρέπει να περιμένουμε μέχρι να η ανάρτηση αυτή να σταματήσει να συγκεντρώνει νέες εντυπώσεις. Σίγουρα τα οι τιμές που συγκεντρώνουμε στην περίοδο που διαφημίζεται μία ανάρτηση δίνουν μία καλή εικόνα για το πόσο καλά ή όχι θα πάει η διαφημιστική μας καμπάνια, αλλά χρειάζεται να εξετάσουμε και τις επόμενες μέρες μετέπειτα για να έχουμε μια πλήρης εικόνα.

Συνοψίζοντας, δημοσιεύοντας την δεύτερη ανάρτηση στοχευμένα στο target group που μας πρότεινε η εφαρμογή, είχαμε ως αποτέλεσμα σε μικρότερο χρονικό διάστημα να επιτύχουμε καλύτερα αποτελέσματα. Είχαμε περισσότερους νέους ακόλουθους συγκριτικά με την πρώτη διαφημίζόμενη ανάρτηση και σχεδόν 500 περισσότερες εντυπώσεις στην ανάρτηση, δηλαδή 500 περισσότερα άτομα είδαν την δημοσίευση μας.

Επίλογος

Ο τομέας της ανάλυσης δεδομένων και η παρουσίαση αυτών ασκεί όλο και μεγαλύτερη επιρροή στη διαδικασία λήψης αποφάσεων στις αποφάσεις που σχετίζονται με το μάρκετινγκ και τη διαφήμιση. Πλέον είναι σημαντικός παράγοντας για την ανάπτυξη κάθε εταιρείας, η δυνατότητα σωστής ανάλυσης δεδομένων και εφαρμογής στην διαφήμιση.

Η ανάκτηση πληροφοριών για τις καταναλωτικές συμπεριφορές και τις συνήθειες των πελατών παίζει καθοριστικό ρόλο στη δημιουργία target groups. Η ανάλυση πληροφοριών όπως ικανοποίηση, προτιμήσεις και μέση κατανάλωση πελατών βοηθούν στην δημιουργία στοχευμένης διαφημιστικής καμπάνιας για κάθε αγοραστικό κοινό. Ουσιαστικά οι εταιρίες διαφημίζουν στο σωστό κοινό το σωστό προϊόν. Αυτό έχει ως αποτέλεσμα οι εταιρίες να ξοδεύουν λιγότερα χρήματα και να αυξάνουν την αποδοτικότητα της καμπάνιας.

Τα αποτυπώματα των καταναλωτών στα κοινωνικά δίκτυα είναι ιδιαίτερα σημαντικά στην εποχή του εξατομικευμένου μάρκετινγκ. Με αυτές τις πληροφορίες, οι εταιρίες μπορούν να στοχεύουν σε διαδικτυακές κοινότητες, ώστε να κατανοήσουν και να εντοπίσουν μοτίβα στις καταναλωτικές συμπεριφορές, κίνητρα και τάσεις.

To repository της εφαρμογής: <https://github.com/gioandreou/thesis-auth>

Βιβλιογραφία

- [1]. El.wikipedia, Μέσα κοινωνικής δικτύωσης, 2018.
- [2]. Zephoria, The Top 20 Valuable Facebook Statistics, ανασύρθηκε στις 13 Ιουλ. 2018 από: <https://zephoria.com/top-15-valuable-facebook-statistics/>
- [3]. Goldbaum, D., Follow the Leader: Simulations on a Dynamic Social Network, Working Paper Series 155, Finance Discipline Group, UTS Business School, University of Technology, Sydney, 2008
- [4]. Kuan-Yu Lin, Hsi-peng Lu. Why people use social networking sites: An empirical study integrating network externalities and motivation theory, Computers in Human Behavior, Science Direct, Pages 1152-1161, 2011
- [5]. Lai, L.S.L. & Turban, Groups Formation and Operations in the Web 2.0 Environment and Social Networks, Group Decision and Negotiation, Springer Link, Pages 387–402, 2008
- [6]. Dube, R., Characteristics of Social Networks, ανασύρθηκε στις 16 Ιουλ. 2018 από:
https://socialnetworking.lovetoknow.com/Characteristics_of_Social_Networks
- [7]. Desreumaux, G., The 10 Top Reasons Why We Use Social ανασύρθηκε στις 16 Ιουλ. 2018 από: <https://wersm.com/the-10-top-reasons-why-we-use-social-networks/>
- [8]. El.wikipedia, Facebook, 2018
- [9]. Moreau, E., The Top Social Networks People Are Using Today, ανασύρθηκε στις 16 Ιουλ. 2018 από: <https://www.lifewire.com/top-social-networking-sites-people-are-using-3486554>
- [10]. Campbell, S., How Does Facebook Work, ανασύρθηκε στις 16 Ιουλ. 2018 από: <https://www.makeuseof.com/tag/facebook-work-nuts-bolts-technology-explained/>
- [11]. El.wikipedia, LinkedIn, 2018
- [12]. El.wikipedia, Google+, 2018

- [13]. Shah, S, The history of social networking, ανασύρθηκε στις 18 Ιουλ. 2018 από: <https://www.digitaltrends.com/features/the-history-of-social-networking/>
- [14]. Perrin, A., Social Media Usage: 2005-2015, Pew Research Center, 2015
- [15]. En.wikipedia, Social media marketing, 2018
- [16]. M. Saravanakumar, T.SuganthaLakshmi, Social Media Marketing, Life Science Journal, Pages 4444-4451, 2012
- [17]. El.wikipedia, Δεδομένα, 2018
- [18]. En.wikipedia, Data, 2018
- [19]. En.wikipedia, Big data, 2018
- [20]. Sas.com, What is Big Data and why it matters, ανασύρθηκε στις 17 Σεπ. 2018 από: https://www.sas.com/en_us/insights/big-data/what-is-big-data.html
- [21]. Μαργαρίτης, Δ., Καρκατζούνης, Β., Σιούφας, Μ., Ιγγλεζάκης, Ι., GDPR ανασύρθηκε στις 17 Σεπ. 2018 από: https://www.lawspot.gr/gdpr?lspt_context=gdpr
- [22]. ΠΛΑΤΗΣ, Ε., ΠΡΟΣΩΠΙΚΑ ΔΕΔΟΜΕΝΑ (2018), ΕΚΔΟΣΕΙΣ ΠΑΠΑΔΟΠΟΥΛΟΣ, Pages 1-43.
- [23]. Sidath Asiri, Data Mining in Brief, ανασύρθηκε στις 19 Σεπ. 2018 από: <https://towardsdatascience.com/data-mining-in-brief-26483437f178>
- [24]. En.wikipedia, Data mining, 2018
- [25]. En.wikipedia, Social media mining, 2018
- [26]. Hill, K., Targeted social media marketing, ανασύρθηκε στις 23 Σεπ. 2018 από: <https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/-3397f7a86668>
- [27]. Richards, Neil M. and King, Jonathan, Three Paradoxes of Big Data, 66 Stanford Law Review Online 41, 2013
- [28]. En.wikipedia, Application programming interface, 2018
- [29]. En.wikipedia, Data analysis, 2018

- [30]. Shelby, B., Exploratory and Confirmatory Analysis: What's the Difference ?, ανασύρθηκε στις 3 Οκτ. 2018 από:
<https://www.sisense.com/blog/exploratory-confirmatory-analysis-whats-difference/>
- [31]. Sas.com, Machine Learning: What it is and why it matters, ανασύρθηκε στις 3 Οκτ. 2018 από: https://www.sas.com/en_us/insights/analytics/machine-learning.html
- [32]. Dr. Klaus Manhart, Was Sie über Maschinelles Lernen wissen müssen, ανασύρθηκε στις 10 Οκτ. 2018 από: <https://www.computerwoche.de/a/was-sie-ueber-maschinelles-lernen-wissen-muessen,3329560>
- [33]. En.wikipedia, Machine learning, 2018
- [34]. Papuc, I., An Introduction to Machine Learning Theory and Its Applications: A Visual Tutorial with Examples, ανασύρθηκε στις 10 Οκτ. 2018 από:
<https://www.datasciencecentral.com/profiles/blogs/an-introduction-to-machine-learning-theory-and-its-applications-a>
- [35]. Vishal, M., Machine Learning for Humans, Part 3: Unsupervised Learning, ανασύρθηκε στις 10 Οκτ. 2018 από: <https://medium.com/machine-learning-for-humans/unsupervised-learning-f45587588294>
- [36]. Devin, S., Supervised vs. Unsupervised, ανασύρθηκε στις 10 Οκτ. 2018 από:
<https://towardsdatascience.com/supervised-vs-unsupervised-learning-14f68e32ea8d>
- [37]. Brownlee, J., Supervised and Unsupervised Machine Learning Algorithms, ανασύρθηκε στις 10 Οκτ. 2018 από:
<https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/>
- [38]. El.wikipedia, Ελληνική Στατιστική Αρχή, 2018