



ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ

Π.Μ.Σ. ΣΤΗ ΛΟΓΙΣΤΙΚΗ ΚΑΙ ΧΡΗΜΑΤΟΟΙΚΟΝΟΜΙΚΗ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Τεχνικές Ανάλυσης Μεγάλων Δεδομένων με την Χρήση Υπολογιστικών Φύλλων (Excel)



ΚΟΥΛΑΛΗΣ ΝΙΚΟΛΑΟΣ

Επιβλέπων Καθηγητής: Πρωτόγερος Νικόλαος

Θεσσαλονίκη, Μάρτιος 2023

Περιεχόμενα

.....	1
Περίληψη	5
Abstract	5
Κεφάλαιο 1 ^ο : Εισαγωγή	6
Κεφάλαιο 2 ^ο : Βιβλιογραφική Ανασκόπηση.....	7
Μέρος Α:.....	7
2.1. Δεδομένα.....	7
2.2. Τα βήματα επεξεργασίας δεδομένων ως το τελικό αποτέλεσμα	8
2.2.1. Συλλογή Δεδομένων.....	8
2.2.2 Διαχείριση Δεδομένων.....	8
2.2.3. Εξόρυξη Δεδομένων	9
2.2.4. Μοντελοποίηση Δεδομένων.....	10
2.2.5. Οπτικοποίηση Δεδομένων	10
2.2.6. Ηλεκτρονική Επεξεργασία Δεδομένων	11
2.2.7. Διατήρηση Δεδομένων.....	11
2.3. Τι είναι τα Big Data;	12
2.3.1. Ιστορική αναδρομή.....	13
2.4. Χαρακτηριστικά των Big Data	14
2.4.1. Όγκος (Volume).....	14
2.4.2. Ταχύτητα (Velocity).....	15
2.4.3. Ποικιλία ή ποικιλομορφία (Variety).....	15
2.4.4. Ορθότητα (Veracity).....	16
2.4.5. Αξία (Value).....	16
Μέρος Β:.....	17
2.5. Ανάλυση Δεδομένων & Excel	19
2.5.1. Πλεονεκτήματα Excel	20
2.6. Βασικές Λειτουργίες Ανάλυσης Δεδομένων του Excel	21
2.6.1. Data entry	22
2.6.2. CONCATENATE.....	24
2.6.3. LEN.....	24
2.6.4. NETWORKDAYS	25
2.6.5. COUNTIF / COUNTIFS	26

2.6.6. COUNTA.....	27
2.6.7. SUMIF / SUMIFS	27
2.6.8. IF.....	28
2.6.9. VLOOKUP & HLOOKUP	29
2.6.10. PivotTables & PivotCharts	31
2.7. Μέθοδοι – Τεχνικές ανάλυσης δεδομένων	38
2.7.1. Σειρές και Πίνακες και υποσύνολα με εύρη.....	38
2.7.2. Καθαρισμός Δεδομένων	39
2.7.3. Μορφοποίηση υπό όρους.....	40
2.7.4. Ταξινόμηση και φιλτράρισμα.....	40
.....	43
2.7.5. Οπτικοποίηση Δεδομένων	43
2.7.6. Εργασία με πολλαπλά φύλλα εργασίας.....	45
2.7.7. What if Analysis	45
2.8. Συμπεράσματα Κεφαλαίου	46
Κεφάλαιο 3: Μεθοδολογία	47
Κεφάλαιο 4: Εφαρμογή του Excel για ανάλυση δεδομένων	50
4.1. Εφαρμογή 1.....	50
4.2. Εφαρμογή 2.....	54
4.3. Συνδυαστικά χαρακτηριστικά των δύο εφαρμογών	57
4.3.1. Τεχνικά Χαρακτηριστικά.....	58
4.3.2. Ποιοτικά Χαρακτηριστικά	60
Κεφάλαιο 5: Συμπεράσματα	61
ΒΙΒΛΙΟΓΡΑΦΙΑ	65

ΔΗΛΩΣΗ ΤΗΡΗΣΗΣ ΑΚΑΔΗΜΑΪΚΗΣ ΔΕΟΝΤΟΛΟΓΙΑΣ

Με πλήρη επίγνωση των επιπτώσεων του νόμου περί πνευματικών δικαιωμάτων, δηλώνω εγγράφως ότι είμαι βοήθεια ο αποκλειστικός συγγραφέας αυτής της Διπλωματικής Εργασίας και ότι κάθε που έλαβα αναφέρεται πλήρως στην εργασία αυτή με λεπτομέρεια. Έχω αναφέρει όλες τις πηγές χρήσης δεδομένων, απόψεων, θέσεων και προτάσεων, ιδεών και λεκτικών αναφορών πλήρως και με σαφήνεια, είτε κατά κυριολεξία είτε βάσει επιστημονικής παράφρασης. Αναλαμβάνω προσωπική και ατομική ευθύνη για την επαρκή τήρηση όλων των δηλώσεων που έχω κάνει και των παραπάνω στοιχείων, και αν εντοπίσω παραβάσεις, είμαι υπόλογος για πνευματική κλοπή. Αυτό μπορεί να γίνει στην αποτυχία της Διπλωματικής μου Εργασίας, καθώς και σε άλλες συνέπειες με τον νόμο περί πνευματικών δικαιωμάτων. Επομένως, δηλώνω ρητά ότι αυτή η Διπλωματική Εργασία είναι αποτέλεσμα της δικής μου προσωπικής προσπάθειας.

Περίληψη

Η παρούσα Διπλωματική Εργασία έχει ως αντικείμενο την μελέτη των Υπολογιστικών Φύλλων και συγκεκριμένα του Excel, ως προς την ικανότητά τους στην Ανάλυση των Μεγάλων Δεδομένων.

Σκοπός είναι να εξάγει αποτελέσματα για την καταλληλότητα και την αξιοπιστία χρήσης του συγκεκριμένου εργαλείου. Η ανάλυση που γίνεται στα ακόλουθα παραδείγματα είναι και αυτή που θα μας δώσει τις απαντήσεις στο ερώτημα που τίθεται. Παράλληλα, πέρα από τα χαρακτηριστικά του Excel, θα καταλήξουμε και συμπεράσματα που αφορούν τα χαρακτηριστικά του ηλεκτρονικού υπολογιστή, τις απαιτήσεις που υπάρχουν και τέλος, στον καλύτερο δυνατό τρόπο ανάλυσης των Μεγάλων Δεδομένων, που ολοένα και περισσότερο τα συναντάμε στην καθημερινότητά μας, γεγονός που συμβαδίζει με την εξέλιξη της τεχνολογίας και των απαιτήσεων από αυτήν.

Abstract

The subject of this Diploma Thesis is the study of Spreadsheets and specifically Excel, in terms of their ability to analyze Big Data.

The purpose is to extract results for the appropriateness and reliability of use of the specific tool. The analysis carried out in the following examples is also the one that will give us the answers to the question posed. At the same time, beyond the features of Excel, we will also come to conclusions regarding the features of the computer, the requirements that exist and finally, the best possible way to analyze Big Data, which we increasingly encounter in our daily lives, a fact that goes hand in hand with the evolution of technology and its requirements.

Κεφάλαιο 1^ο: Εισαγωγή

Στην σημερινή εποχή, ο κόσμος έχει μετατραπεί σε κοινωνία της πληροφορίας, με τα Μεγάλα Δεδομένα να βρίσκονται σε όλους τους τομείς της παγκόσμιας οικονομίας. Η συνεχιζόμενη αύξηση του όγκου δεδομένων είναι στην πραγματικότητα απεριόριστη. Η ανάπτυξη της τεχνολογίας με τόσο γρήγορους ρυθμούς την καθιστά σε θέση να χειριστεί τις ποσότητες από δομημένα και αδόμητα δεδομένα από διαφορετικές πηγές, ανοίγοντας νέους δρόμους έρευνας και εφαρμογών, με άμεση επιρροή σε όλους τους τομείς της κοινωνίας μας.

Η συλλογή και η επεξεργασία κάθε είδους δεδομένων με σκοπό την χρήση τους σε διάφορες εφαρμογές πραγματοποιείται τόσο από ιδιωτικούς όσο και από δημοσίους φορείς. Η εφαρμογή των big data στοχεύει στην κατόπιν κατάλληλης επεξεργασίας εξαγωγή της ή των πληροφοριών που κρίνονται ουσιώδους σημασίας κατά την επιδίωξη ενός συγκεκριμένου σκοπού. Για την υλοποίηση της ανάλυσής τους ολοένα και περισσότερα λογισμικά αναπτύσσονται στην αγορά καθημερινά με τον ανταγωνισμό να φτάνει σε πολύ υψηλά επίπεδα.

Αντικείμενο της παρούσας εργασίας αποτελεί η διερεύνηση της καταλληλότητας των υπολογιστικών φύλλων και συγκεκριμένα του excel για την ανάλυση των Μεγάλων Δεδομένων. Αυτό αποτελεί και το ερευνητικό ερώτημα που θα προσπαθήσει να απαντηθεί μέσω της βιβλιογραφικής ανασκόπησης και έπειτα με την εφαρμογή της μεθοδολογίας.

Στο επόμενο κεφάλαιο θα επιχειρηθεί ο προσδιορισμός σημαντικών και πολύ βασικών εννοιών για την ανάλυση δεδομένων με σκοπό να κάνει τον αναγνώστη οικείο με τον θέμα της εργασίας. Παράλληλα στο δεύτερο μέρος, θα γίνει αναλυτική απόδοση της χρήσης των σημαντικότερων εντολών που χρησιμοποιούμε στο excel για την ανάλυση των δεδομένων, σε συνδυασμό με οπτικές αναφορές για την ευκολότερη κατανόηση της χρήσης τους. Στην συνέχεια, στο τρίτο κεφάλαιο θα αναλυθεί ο τρόπος με τον οποίο θα καταλήξουμε στην απάντηση του ερευνητικού ερωτήματος που τέθηκε στην παρούσα ενότητα. Το τέταρτο κεφάλαιο όπου αφορά την ανάλυση, θα αποτελέσει την πρακτική εφαρμογή όλων των παραπάνω, πάνω σε πραγματικά δεδομένων με σκοπό να μελετηθεί η λειτουργία των εντολών του excel και η ανταπόκρισή του ανάλογα με την κάθε περίπτωση και τις διαφοροποιήσεις που μπορεί να συναντήσουμε. Τέλος, καταλήγουμε στα συμπεράσματα και παράλληλα την απάντηση του ερευνητικού μας ερωτήματος.

Κεφάλαιο 2^ο: Βιβλιογραφική Ανασκόπηση

Μέρος Α:

Σκοπός του συγκεκριμένου κεφαλαίου είναι να δώσει την ερμηνεία σημαντικών εννοιών οι οποίες θα αποτελέσουν την βάση για την κατανόηση των τεχνικών και των εργαλείων που θα χρησιμοποιηθούν στα επόμενα κεφάλαια. Παράλληλα γίνεται εκτενής αναφορά στον τρόπο λειτουργίας όλων των απαραίτητων εντολών του excel για να γίνει στην πορεία, εύκολα κατανοητή η εφαρμογή τους από οποιονδήποτε αναγνώστη της παρούσας Διπλωματικής Εργασίας.

2.1. Δεδομένα

Τα δεδομένα αποτελούν μεμονωμένα γεγονότα, τα οποία μπορεί να είναι στατιστικά στοιχεία ή στοιχεία πληροφοριών και συχνά είναι αριθμητικού τύπου. Αν χρησιμοποιήσουμε μια έννοια η οποία είναι πιο τεχνική, τα δεδομένα αντιπροσωπεύουν ένα σύνολο τιμών οι οποίες μπορεί να είναι είτε ποιοτικές είτε ποσοτικές μεταβλητές για ένα ή περισσότερα πρόσωπα ή αντικείμενα, ενώ ένα δεδομένο είναι μια ενιαία τιμή μιας μεμονωμένης μεταβλητής (Στεφανιδάκης, Ανδρόνικος, & Παπαδάκης, 2015).

Συχνά οι όροι δεδομένα και πληροφορίες χρησιμοποιούνται με την ίδια έννοια, όμως στην πραγματικότητα πρόκειται για δύο ξεχωριστές έννοιες. Σε αρκετές δημοσιεύσεις λέγεται πως τα δεδομένα μετατρέπονται σε πληροφορίες όταν αυτά εξετάζονται. Ωστόσο, στις ακαδημαϊκές επεξεργασίες του υποκειμένου τα δεδομένα είναι απλώς μονάδες πληροφοριών. Τα δεδομένα χρησιμοποιούνται στην επιστημονική έρευνα, στη διαχείριση των επιχειρήσεων (δεδομένων πωλήσεων, εσόδων, κερδών), στη διακυβέρνηση (ποσοστά εγκληματικότητας / ανεργίας) και ουσιαστικά σε κάθε άλλη μορφή ανθρώπινης οργανωτικής δραστηριότητας. (Sodhar, Jalban, Buller & Naz, 2022).

Γενικότερα τα δεδομένα συνεισφέρουν στην λήψη των αποφάσεων, καθώς αποτελούν μικρότερες μονάδες από πραγματικές πληροφορίες οι οποίες μπορούν να χρησιμοποιηθούν ως βάση για συζήτηση και υπολογισμούς. Τα δεδομένα μετρώνται, συλλέγονται, αναλύονται και χρησιμοποιούνται για την δημιουργία οπτικοποιημένων γραφημάτων, πινάκων ή εικόνων. Τα ακατέργαστα δεδομένα αποτελούν μια συλλογή χαρακτήρων ή αριθμών πριν αυτή «καθαριστεί» και πάρει την επιθυμητή μορφή από τους ερευνητές. Τα ανεπεξέργαστα δεδομένα πρέπει να διορθωθούν και να αφαιρεθούν τυχόν ακραίες τιμές ή σφάλματα όπως για παράδειγμα είναι μια λανθασμένη ένδειξη θερμομέτρου. Η διαδικασία της επεξεργασίας δεδομένων πραγματοποιείται συνήθως κατά στάδια και τα «επεξεργασμένα δεδομένα» περνάνε στην φάση όπου πλέον, μπορούν να θεωρηθούν τα «ακατέργαστα δεδομένα» του επόμενου σταδίου. (Yang, Yan, & Li, 2022).

2.2. Τα βήματα επεξεργασίας δεδομένων ως το τελικό αποτέλεσμα

Σχεδόν όλοι οι οργανισμοί και οι επιχειρήσεις που δραστηριοποιούνται στο διαδίκτυο και όχι μόνο, επεξεργάζονται δεδομένα με έναν απλό ή με πιο σύνθετους τρόπους, ανάλογα με τους στόχους που επιδιώκουν να πετύχουν. Στην ουσία με τον όρο επεξεργασία δεδομένων, αναφερόμαστε στην συλλογή και χειραγώγηση των δεδομένων που μας επιτρέπουν την παραγωγή σχετικών πληροφοριών. Τα ακατέργαστα δεδομένα δεν παρέχουν σωστή πληροφόρηση και επομένως πρέπει να συγκεντρωθούν, να εμπλουτιστούν, να φιλτραριστούν και να καθαριστούν για να φτάσουμε στο τελικό αποτέλεσμα. Παρακάτω στην παρούσα ενότητα, παρουσιάζονται τα βήματα που πρέπει να ακολουθηθούν έτσι ώστε να πραγματοποιηθεί σωστά και μεθοδικά η επεξεργασία των δεδομένων.

2.2.1. Συλλογή Δεδομένων

Η συλλογή των δεδομένων μπορεί να επιτευχθεί με δύο τρόπους, τον πρωτογενή και τον δευτερογενή. Στην περίπτωση του πρώτου ο ερευνητής θα είναι ο πρώτος που θα λάβει «στα χέρια του» τα δεδομένα ενώ στην περίπτωση του δεύτερου η συλλογή αυτών έχει ήδη πραγματοποιηθεί από μια δευτερεύουσα πηγή. Το πλήθος των μεθοδολογιών ανάλυσης δεδομένων είναι ευρύ και αυτό περιλαμβάνει τριγωνισμό δεδομένων και διήθηση δεδομένων. Η διήθηση δεδομένων προσφέρει μια αρθρωτή μέθοδο συλλογής, ταξινόμησης και ανάλυσης δεδομένων χρησιμοποιώντας πέντε πιθανές γωνίες ανάλυσης (τουλάχιστον τρεις) που έχει ως σκοπό να μεγιστοποιηθεί η αντικειμενικότητα της έρευνας και να επιτραπεί η όσο το δυνατόν πληρέστερη κατανόηση των υπό διερεύνηση φαινομένων: ποιοτικές και ποσοτικές μέθοδοι, βιβλιογραφικές ανασκοπήσεις (συμπεριλαμβανομένων επιστημονικών άρθρων), συνεντεύξεις με ειδικούς και προσομοίωση υπολογιστή. Στη συνέχεια, τα δεδομένα εισέρχονται χρησιμοποιώντας μια σειρά προκαθορισμένων βημάτων έτσι ώστε να εξαχθούν οι πιο σχετικές πληροφορίες. (Γιόγος Α., Μπούζας Α., 2022)

2.2.2 Διαχείριση Δεδομένων

Στη σύγχρονη χρήση της διαχείρισης δεδομένων, βλέπουμε τον όρο δεδομένα να αντικαθίσταται ολοένα και περισσότερο από πληροφορίες ή ακόμα και από τον όρο γνώση σε ένα μη τεχνικό πλαίσιο. Συνεπώς η διαχείριση δεδομένων καταλήγει να γίνεται διαχείριση πληροφοριών ή διαχείριση γνώσης. Η διαχωριστική γραμμή ανάμεσα στα δεδομένα και στην παραγωγική τους αξιοποίηση απεικονίζεται μέσω της σκάλας πληροφοριών. Ωστόσο, με την πάροδο του χρόνου, η ανάδυση της έννοιας των Big Data έφερε μια επανεκτίμηση των δεδομένων, αφού αναφέρεται στη συλλογή και την ανάλυση των τεράστιων δεδομένων. (Στεφανιδάκης, Ανδρόνικος, & Παπαδάκης, 2015).

Αρκετοί οργανισμοί έχουν δημιουργήσει δικά τους κέντρα διαχείρισης δεδομένων (Data Management Centers / DMC) για τις δραστηριότητές τους. Αυτά αποτελούνται

από διάφορα εργαλεία τα οποία κάνουν ευκολότερη την διαχείριση των δεδομένων και βελτιώνουν την απόδοση. Μερικά παραδείγματα είναι:

- Δημιουργία εφαρμογών.
- Ενίσχυση της συνεργασίας των προγραμματιστών με τους αρχιτέκτονες.
- Αυτοματοποίηση και απλοποίηση λειτουργιών.

2.2.3. Εξόρυξη Δεδομένων

Η εξόρυξη δεδομένων (data mining) είναι μια νέα δυναμική τεχνολογία που βοηθάει τις επιχειρήσεις να εστιάσουν στην σημαντική πληροφορία που βρίσκεται μέσα στις αποθήκες δεδομένων τους (data warehouses). Οι τεχνικές της είναι σε θέση να αναζητήσουν και να βρουν γρήγορα και λεπτομερειακά βάσεις δεδομένων για την αναζήτηση κρυμμένων προτύπων. Σύμφωνα με τα παραπάνω μπορούμε να πούμε ότι η εξόρυξη γνώσης είναι μια διαδικασία εξαγωγής κρυμμένης πληροφορίας από μεγάλες βάσεις δεδομένων.

Ο συγκεκριμένος όρος της εξόρυξης δεδομένων αποτελεί μια λανθασμένη ονομασία, καθώς ο σκοπός αυτά τα δεδομένα να προσφέρουν εξαγωγή προτύπων και γνώσης από τις πηγές μεγάλων δεδομένων και όχι η εξόρυξη των ίδιων των δεδομένων αυτών καθ' αυτών. Επίσης, συχνά εφαρμόζεται σε οποιαδήποτε μορφή επεξεργασίας δεδομένων ή πληροφοριών μεγάλης κλίμακας (συλλογή, εξαγωγή, αποθήκευση, ανάλυση και στατιστικά στοιχεία) καθώς και σε οποιαδήποτε εφαρμογή συστήματος υποστήριξης αποφάσεων υπολογιστή, οι οποίες συμπεριλαμβάνουν και την τεχνητή νοημοσύνη. Συχνά συναντάμε πιο γενικούς όρους όπως ανάλυση και ανάλυση δεδομένων οι οποίοι σε περιπτώσεις όπου αναφέρονται σε πραγματικές μεθόδους, όπως η τεχνητή νοημοσύνη και η μηχανική μάθηση είναι πιο κατάλληλοι.

Η διαφορά μεταξύ ανάλυσης δεδομένων και εξόρυξης δεδομένων είναι ότι η ανάλυση δεδομένων χρησιμοποιείται για τη δοκιμή μοντέλων και υποθέσεων σχετικά με το σύνολο δεδομένων, όπως για παράδειγμα για την ανάλυση της αποτελεσματικότητας μιας διαδικασίας διαφήμισης, ανεξάρτητα από τον όγκο των δεδομένων που θα χρειαστούν. Αντίθετα, η εξόρυξη δεδομένων χρησιμοποιεί μηχανική μάθηση και στατιστικά μοντέλα για να αποκαλύψει μυστικά ή κρυφά μοτίβα σε μεγάλο όγκο δεδομένων. (Τσάκαλης, 2020).

Η εξόρυξη των δεδομένων συμπεριλαμβάνει 6 κατηγορίες εργασιών:

- Ανίχνευση ανωμαλιών. – Αναφέρεται στον εντοπισμό ασυνήθιστων εγγραφών ή σφαλμάτων που χρειάζονται περαιτέρω διερεύνηση.
- Εκμάθηση κανόνων συσχέτισης. – Αφορά την σύγκριση μεταξύ μεταβλητών. Για παράδειγμα μία εταιρία μια εταιρία κινητής τηλεφωνίας μπορεί να συγκεντρώσει στοιχεία σχετικά με τις πωλήσεις δύο κινητών μοντέλων της και να χρησιμοποιήσει τις πληροφορίες αυτές για σκοπούς μάρκετινγκ.
- Ομαδοποίηση. – Αποτελεί το έργο δημιουργίας ομάδων και δομών για τα δεδομένα που κατά κάποιο τρόπο μπορούν να χαρακτηριστούν ως παρόμοια.

- Ταξινόμηση. – Έχει ως καθήκον την γενίκευση της γνωστής δομής για εφαρμογή σε νέα δεδομένα. Ένα παράδειγμα αποτελεί ένα πρόγραμμα ηλεκτρονικού ταχυδρομείου που έχει την δυνατότητα να ταξινομήσει ένα εισερχόμενο μήνυμα ως ανεπιθύμητο.
- Παλινδρόμηση. – Επιχειρεί μέσω αυτής να βρει μια συνάρτηση που μοντελοποιεί τα δεδομένα με το λιγότερο δυνατόν σφάλμα δηλαδή, με σκοπό την εκτίμηση των σχέσεων μεταξύ δεδομένων ή συνόλων δεδομένων.
- Σύνοψη – παρέχει στο κοινό μια πιο συμπαγή αναπαράσταση του συνόλου δεδομένων, συμπεριλαμβανομένης και της απεικόνισης και της δημιουργίας αναφορών.

2.2.4. Μοντελοποίηση Δεδομένων

Η μοντελοποίηση των δεδομένων για την μηχανική λογισμικού αποτελεί την διαδικασία δημιουργίας ενός μοντέλου δεδομένων για ένα πληροφοριακό σύστημα με την εφαρμογή ορισμένων επίσημων τεχνικών. (Shaw, 2022) Αποτελεί μια διαδικασία που χρησιμοποιείται για την ανάλυση και τον καθορισμό των απαιτήσεων δεδομένων τα οποία είναι απαραίτητα για την επίτευξη της υποστήριξης των επιχειρηματικών διαδικασιών των πληροφοριακών συστημάτων των επιχειρήσεων. Για αυτό τον λόγο η παραπάνω διαδικασία απαιτεί επαγγελματίες διαμορφωτές δεδομένων σε συνεργασία με το προσωπικό των επιχειρήσεων αλλά και με τους χρήστες αυτών.

Υπάρχουν 3 διαφορετικοί τύποι μοντέλων δεδομένων που παράγονται σταδιακά μέχρι να καταλήξουν στην πραγματική βάση δεδομένων η οποία είναι και αυτή που θα χρησιμοποιηθεί για το πληροφοριακό σύστημα. Αρχικά τα δεδομένα ως ένα σύνολο προδιαγραφών και χρησιμοποιούνται για την συζήτηση με τα ενδιαφερόμενα μέρη της επιχείρησης. Στην συνέχεια το εννοιολογικό μοντέλο μεταφράζεται σε ένα λογικό μοντέλο δεδομένων και τεκμηριώνει δομές δεδομένων που μπορούν να εφαρμοστούν σε βάσεις δεδομένων. Το τελευταίο βήμα της μοντελοποίησης είναι η απόδοση των λογικών δεδομένων σε μοντέλο φυσικών δεδομένων το οποίο οργανώνει τα δεδομένα σε πίνακες και στην συνέχεια είναι αυτό που υπολογίζει τις λεπτομέρειες πρόσβασης, απόδοσης και αποθήκευσης. Η μοντελοποίηση δεν περιορίζεται μόνο στο να ορίζει στοιχεία δεδομένων, αλλά επεκτείνεται και τις δομές τους και τις σχέσεις που υφίστανται μεταξύ τους (Ashraff, 2022).

2.2.5. Οπτικοποίηση Δεδομένων

Η οπτικοποίηση δεδομένων που αναφέρεται στον τρόπο με τον οποίο παρουσιάζονται γραφικά δεδομένα και πληροφορίες, και είναι ένα διεπιστημονικό πεδίο που συνδυάζει τα πεδία της ψυχολογίας, της επικοινωνίας και της πληροφορικής. Αποτελεί έναν αποτελεσματικό τρόπο επικοινωνίας όταν πρέπει να αναπαρασταθούν πολλά δεδομένα και πληροφορίες.

Η οπτικοποίηση δεδομένων βασίζεται στην υπόθεση ότι οι αναπαραστάσεις και οι τεχνικές αλληλεπίδρασης που γνωστοποιούν το ευρύ φάσμα του ανθρώπινου οράματος, έτσι ώστε οι χρήστες να μην μπορούν να βλέπουν, να εξερευνούν και να κατανοούν μεγάλες ποσότητες πληροφοριών ταυτόχρονα. Η οπτικοποίηση πληροφοριών επικεντρώνεται στη δημιουργία αφηρημένων αναπαραστάσεων, που

παρουσιάζονται με εύκολα αντιληπτούς τρόπους. (Στεφανιδάκης, Ανδρόνικος, & Παπαδάκης, 2015).

Η ανάλυση δεδομένων αποτελεί ένα τμήμα της εφαρμοσμένης έρευνας και της επίλυσης προβλημάτων. Υπάρχουν πολλές βασικές προσεγγίσεις ανάλυσης δεδομένων, όπως η οπτικοποίηση μέσω τύπων γραφημάτων (ιστογράμματα, διαγράμματα διασποράς, γραφήματα παράλληλων συντεταγμένων), οι μέθοδοι μηχανικής μάθησης (ομαδοποίηση, ταξινόμηση), η εξόρυξη δεδομένων και τα στατιστικά μοντέλα (παλινδρόμηση, δοκιμές υποθέσεων). Ειδικότερα, η ανάλυση των οπτικών δεδομένων απαιτεί γνωστικές δεξιότητες και δημιουργικότητα από τους αναλυτές, καθώς η ορθή ερμηνεία των γραφημάτων μπορεί να προσφέρει πολύτιμη πληροφορία. (Γιόγος Α., & Μπούζας Αντώνιος, 2022).

2.2.6. Ηλεκτρονική Επεξεργασία Δεδομένων

Η αυτοματοποίηση είναι μια διαδικασία που χρησιμοποιείται στην ηλεκτρονική επεξεργασία δεδομένων για την απλοποίηση και την επιτάχυνση επαναλαμβανόμενων διαδικασιών. Αυτό συμβαίνει συνήθως σε περιπτώσεις που χρειάζεται η επεξεργασία μεγάλου όγκου δεδομένων. Παραδείγματα περιλαμβάνουν την αυτόματη καταγραφή τραπεζικών συναλλαγών και την αυτόματη ενημέρωση των αποθεμάτων μιας εταιρείας. Ο όρος «ηλεκτρονικός» ή «αυτόματος» τροποποιητής χρησιμοποιείται για να διακρίνει τη διαδικασία αυτή από την ανθρώπινη επεξεργασία δεδομένων γραφείου. Στην ουσία, ο στόχος της αυτοματοποίησης είναι η εξοικονόμηση χρόνου και η αύξηση της ακρίβειας και της αποδοτικότητας της επεξεργασίας δεδομένων. (Sodhar, Jalbani, Buller, & Naz, 2022).

2.2.7. Διατήρηση Δεδομένων

Η διατήρηση των δεδομένων τη λήψη μέτρων για την προστασία και την ακεραιότητά τους. Αυτά τα μέτρα περιλαμβάνουν πολιτικές, πολιτικές και στρατηγικές που επιδιώκουν την παράταση της ύπαρξης και της αυθεντικότητας των δεδομένων. Η κύρια σκοπιμότητα της διατήρησης των δεδομένων είναι η προστασία τους από απώλεια ή καταστροφή και η ενθάρρυνση της επαναχρησιμοποίησης και της εξέλιξής τους.

Η απώλεια δεδομένων είναι μία από τις μεγαλύτερες απειλές για την ακεραιότητα και την ασφάλεια των δεδομένων, ανεξάρτητα από το είδος τους και την κλίμακα στην οποία λειτουργούν. Μπορεί να προκληθεί από διάφορους τρόπους, όπως κακόβουλη ενέργεια, τεχνικές βλάβες, φυσικές καταστροφές και αμέλεια. Η απώλεια δεδομένων μπορεί να έχει σοβαρές συνέπειες σε σημαντικούς τομείς όπως η ιατρική, η ασφάλεια, η οικονομία και ο πολιτισμός, και για αυτόν τον λόγο είναι ζωτικής σημασίας η ανάληψη δράσης για την προστασία των δεδομένων. (Yang, Yan, & Li, 2022).

2.3. Τι είναι τα Big Data;

Τα Big Data, ή αλλιώς στην γλώσσα μας «Μεγάλα Δεδομένα», είναι ένας ευρύς όρος που αναφέρεται στον τεράστιο όγκο δεδομένων και απαιτείται η ανάπτυξη νέων εργαλείων και μεθόδων για την ανάλυσή τους. Οι παραδοσιακές τεχνικές και εφαρμογές των ημερών μας καθίστανται ανεπαρκείς, διότι όσο αναπτύσσονται τα Big Data, τόσο περισσότερες προκλήσεις κάνουν την εμφάνισή τους. Μεταξύ των βασικών προκλήσεων είναι η συλλογή, η αποθήκευση, η επεξεργασία και η κοινή χρήση των δεδομένων, η ανάλυσή τους και η απεικόνιση των αποτελεσμάτων που προκύπτουν.

Τα τελευταία χρόνια, η ανάγκη για την διαχείριση των Μεγάλων Δεδομένων ή αλλιώς Big Data, έχει αυξηθεί λόγω των μεγάλων αναγκών που υπάρχουν στην παραγωγή. Η πρώτη εμφάνιση του συγκεκριμένου όρου έγινε το 1997 από τους επιστήμονες της NASA. Συγκεκριμένα ανέφεραν πως υπήρχε αδυναμία στην αναπαράσταση συνόλων δεδομένων γραφικά, καθώς ήταν τόσο πολλά σε όγκο που δεν επαρκούσε η κύρια αποθηκευτική μνήμη του δίσκου. Αυτό ήταν και ένα από τα πρώτα προβλήματα Μεγάλων Δεδομένων.

Η τεχνολογική εξέλιξη έδωσε την δυνατότητα να δημιουργηθούν μηχανισμοί παρακολούθησης των λειτουργιών ενός οργανισμού σε ανεπτυγμένο επίπεδο. Μέσω της ψηφιοποίησης των διαδικασιών παραγωγής, οι εταιρίες και οι οργανισμοί έχουν καταφέρει να παράγουν τεράστιους όγκους δεδομένων σε πολύ σύντομο χρονικό διάστημα. Τα δεδομένα αυτά κρύβουν πολύτιμες και χρήσιμες πληροφορίες για τις επιχειρήσεις και η ανάλυσή τους μπορεί να οδηγήσει την επιχείρηση να λάβει σημαντικές αποφάσεις, όπως για παράδειγμα μπορεί να είναι η βελτιστοποίηση της παραγωγής, η αποφυγή λανθασμένων επενδύσεων, κλπ. Ωστόσο το αποτέλεσμα δεν είναι πάντα το ζητούμενο, καθώς με την λανθασμένη επιλογή τεχνολογικής λύσης.

Οι ανάγκες της εποχής και κατά κύριο λόγο η καθημερινή χρήση του διαδικτύου, οδηγούν τις επιχειρήσεις στο ηλεκτρονικό εμπόριο με αποτέλεσμα η παραγωγή δεδομένων να αυξάνεται με γοργούς ρυθμούς και τα Big Data να βρίσκονται πλέον παντού. Ολοένα και περισσότερος κόσμος, είτε επαγγελματίες είτε καταναλωτές προϊόντων έλκονται από τα Big Data, καθώς μπορούν να τους προσφέρει πολλά οφέλη η αξιοποίησή τους. Στις μέρες μας, η δυνατότητα αποθήκευσης των δεδομένων και η ανάλυση των αποτελεσμάτων, με την βοήθεια της τεχνολογίας, έχει γίνει πιο προσιτή και εφικτή. Παράλληλα, και τα μέσα για την εξόρυξη της γνώσης από τα δεδομένα παρουσιάζουν ταχεία βελτίωση, καθώς τα διαθέσιμα λογισμικά για την εφαρμογή τεχνικών αυξανόμενης πολυπλοκότητας συνδυάζονται με την αυξανόμενη υπολογιστική ισχύ.

Οι ανάγκες της εποχής, οδηγούν τις επιχειρήσεις στο ηλεκτρονικό εμπόριο με αποτέλεσμα η παραγωγή δεδομένων να αυξάνεται με γοργούς ρυθμούς και τα Big Data να βρίσκονται πλέον παντού. Οι επιχειρήσεις πλέον, αποθηκεύουν όλα τους τα δεδομένα σε ψηφιακή μορφή. Αυτό οδηγεί σε νέες προκλήσεις, όπως για παράδειγμα αν είναι διαχειρίσιμος αυτός ο όγκος πληροφορίας. Ωστόσο τα Big Data, δεν παράγονται μόνο από τις επιχειρήσεις αλλά και από το κάθε άτομο ξεχωριστά. Συγκεκριμένα, σχεδόν ο κάθε ένας από εμάς διατηρεί ένα ψηφιακό προφίλ σε ένα ή περισσότερα κοινωνικά δίκτυα, από όπου και παράγονται χρήσιμα δεδομένα. Μεγάλοι

όγκοι δεδομένων παρουσιάζονται ακόμη και σε επιστημονικούς τομείς, όπως η ιατρική, η αστρονομία, η μετεωρολογία, αλλά και η βιολογία, χάρη στις νέες τεχνολογικές καινοτομίες που έχουν βοηθήσει κατά πολύ την ανθρωπότητα. Ο ρυθμός αύξησης των δεδομένων έχει τέτοια αυξητική πορεία που προβλέπεται πως τα επόμενα χρόνια θα υπάρξει μια ακόμη μεγαλύτερη “έκρηξη πληροφορίας” σαν ποτέ άλλοτε.

2.3.1. Ιστορική αναδρομή

Από τα παραπάνω καταλαβαίνουμε πως τα Big Data αποτελούν ένα γρήγορα εξελισσόμενο πεδίο εφαρμογής, το οποίο έχει εισβάλλει σε πολλούς τομείς της σύγχρονης ζωής. Έχουν γίνει αρκετές προσπάθειες να βρεθεί ένας σαφής και πλήρης ορισμός βασισμένος στην επιστημονική έρευνα, κάτι το οποίο όμως δεν είναι εύκολο και ίσως οφείλεται στην ταχεία και ασταμάτητη εξέλιξη αυτών των δεδομένων. Ενδεικτικά αναφέρονται οι παρακάτω ορισμοί:

Ένας από τους πιο διαδεδομένους ορισμούς στη βιβλιογραφία είναι αυτός που συμπεριλήφθηκε σε μια έκθεση της Gartner το 2012: *«Τα μεγάλα δεδομένα είναι πληροφοριακά στοιχεία υψηλού όγκου, υψηλής ταχύτητας και υψηλής ποικιλίας που απαιτούν οικονομικά αποδοτικές, καινοτόμες μορφές επεξεργασίας πληροφοριών για βελτιωμένη ανακάλυψη γνώσεων, λήψη αποφάσεων και βελτιστοποίηση διαδικασιών»* (Gartner IT Glossary 2012).

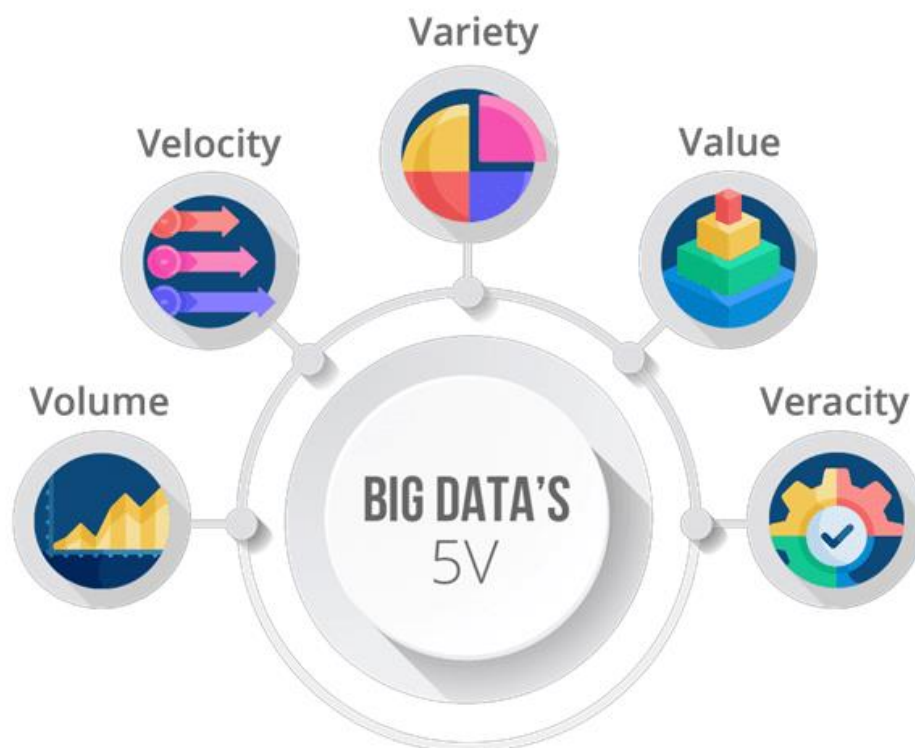
Επίσης, οι Manyika et al. (2011) αναφέρουν τα Big Data ως *«ένα σύνολο στοιχείων και δεδομένων τα οποία δεν μπορούν να συγκεντρωθούν, να αποθηκευτούν και να επεξεργαστούν από τα κλασικά εργαλεία διαχείρισης βάσεων δεδομένων»*.

Τέλος ο ορισμός που δόθηκε από την ομάδα εργασίας του Άρθρου 29, όπου ο εξεταζόμενος όρος *«αναφέρεται στα γιγαντιαία ψηφιακά σύνολα δεδομένων τα οποία αποθηκεύονται από τις επιχειρήσεις, τις κυβερνήσεις και άλλους μεγάλους οργανισμούς και στην συνέχεια αναλύονται σε βάθος με την χρήση υπολογιστικών αλγορίθμων»*.

Οι ορισμοί που έχουν δοθεί είναι δεκάδες. Από την μελέτη στην βιβλιογραφική επισκόπηση σαν γενικό συμπέρασμα, μπορούμε να τα περιγράψουμε σαν συλλογές δεδομένων, τα οποία κατά κύριο λόγο είναι δομημένα σε μεγάλο ποσοστό και ο όγκος τους είναι τέτοιος που καθιστά δύσκολη την αποθήκευση, επεξεργασία και την ανάλυσή τους με την χρήση των παραδοσιακών μεθόδων της επιστήμης της πληροφορικής.

2.4. Χαρακτηριστικά των Big Data

Οι περισσότεροι ορισμοί των “Μεγάλων Δεδομένων” εστιάζουν στον όγκο των δεδομένων. Ωστόσο, εκτός από το μέγεθος των δεδομένων, υπάρχουν και άλλα εξίσου σημαντικά χαρακτηριστικά στα “Μεγάλα Δεδομένα”. Όταν απαιτείται σχεδιασμένη λύση και περιβάλλον ανάλυσης με σκοπό να προκύψουν έγκυρα αποτελέσματα από τα δεδομένα αυτά, πρέπει να υπάρχουν ένα ή περισσότερα από τα παρακάτω χαρακτηριστικά, τα οποία αναφέρονται και ως 5Vs.



Εικόνα 1: 5Vs of Big Data

2.4.1. Όγκος (Volume)

Ο όγκος είναι απόρροια της εκτεταμένης διαδικτυακής δραστηριότητας που λαμβάνει χώρα καθημερινά από τον καθένα μας ξεχωριστά αλλά και από τις επιχειρήσεις. Το γεγονός ότι όλο και περισσότερα μέσα και δραστηριότητες την σήμερα ημέρα, όπως για παράδειγμα το εμπόριο και οι τραπεζικές συναλλαγές μεταφέρονται στον χώρο του διαδικτύου σε συνδυασμό και με τη μείωση του σχετικού κόστους, φέρνει ως αποτέλεσμα την παραγωγή αστρονομικού όγκου δεδομένων ανά δευτερόλεπτο, φαινόμενο που προβλέπεται να συνεχίζει να εντείνεται όλο και περισσότερο κατά τη

διάρκεια των επόμενων χρόνων. Τα οφέλη από τη συλλογή και ανάλυση των παραπάνω δεδομένων, δημιουργούν μια σειρά από προκλήσεις για την απόκτηση πολύτιμης γνώσης για τους ανθρώπους και τις επιχειρήσεις.

2.4.2. Ταχύτητα (Velocity)

Πέρα από τον όγκο τα Big Data χαρακτηρίζονται και από μεγάλη ταχύτητα, εννοώντας ότι τα προς διαχείριση δεδομένα καταφθάνουν στην επιχείρηση και συσσωρεύεται μεγάλος όγκος δεδομένων σε σύντομο χρονικό διάστημα. Μάλιστα σε ορισμένες περιπτώσεις η ταχύτητα αυτή μπορεί να πλησιάζει και την ανταπόκριση σε πραγματικό χρόνο. Μια τέτοια περίπτωση αποτελούν οι αυτοματοποιημένες χρηματιστηριακές συναλλαγές, όπου αυτόματα συστήματα μέσω των αλγορίθμων μπορούν να εκπαιδευτούν να «μελετούν» δεδομένα και βάσει της ανάλυσης στην οποία προβαίνουν, μελλοντικά να λαμβάνουν σχετικές αποφάσεις. Γενικότερα στα big data η έννοια του χρόνου έχει μεγάλη σημασία καθώς όσο πιο πρόσφατη είναι μια πληροφορία τόσο μεγαλύτερη είναι και η αξία αυτής. (Stucke M., Grunes A. 2016)

2.4.3. Ποικιλία ή ποικιλομορφία (Variety)

Η ποικιλία ή ποικιλομορφία αναφέρεται στους πολλαπλούς τύπους δεδομένων αλλά και στις μορφές αυτών. Το συγκεκριμένο χαρακτηριστικό είναι αυτό που προσελκύει διάφορες εταιρίες οι οποίες τα μετατρέπουν, τα επεξεργάζονται, τα αποθηκεύουν και έπειτα τα χρησιμοποιούν για να εξυπηρετήσουν τα συμφέροντά τους. Τα συγκεκριμένα δεδομένα μπορεί να προέρχονται από πολλαπλές πηγές και μέσα και να αφορούν αντίστοιχα πολλούς τομείς ή διαφορετικά αντικείμενα δραστηριότητας. Παραδείγματα πηγών από όπου προέρχονται τα παραπάνω δεδομένα μπορεί να είναι οι άνθρωποι μέσω την αλληλεπίδρασης που υπάρχει με τα ψηφιακά συστήματα και την τεχνολογία αλλά και από τεχνολογικά μέσα, μέσω προγραμμάτων λογισμικού. (Wikipedia.org) Οτιδήποτε από τα παραπάνω και αν ισχύει μπορούν να διακριθούν σε 3 βασικές κατηγορίες, *τα δομημένα και τα αδόμητα* δεδομένα και τα *ημι – δομημένα δεδομένα*.

Τα δομημένα δεδομένα, συχνότερα τα συναντάμε σε μορφή πίνακα και μας δίνουν πληροφορίες για τις σχέσεις μεταξύ κάποιων οντοτήτων σε σχεσιακές βάσεις δεδομένων. Αυτά τα δεδομένα προκύπτουν κυρίως από πληροφοριακά συστήματα όπως τα ERP και CRM.

Τα αδόμητα δεδομένα δεν έχουν συγκεκριμένη μορφή όπως π.χ. κείμενο, εικόνα, ήχος και αυτό οδηγεί στην δημιουργία ειδικών μεθόδων και εργαλείων για να γίνει η επεξεργασία, η ανάλυση αλλά και η αποθήκευσή τους. Ο συγκεκριμένος τύπος δεδομένων είναι και αυτός που τείνει να αυξάνεται ολοένα και περισσότερο.

Τα ημι - δομημένα δεδομένα είναι εκείνα που έχουν δομή αλλά δεν ταιριάζουν στην σχεσιακή βάση δεδομένων. Τα ημι - δομημένα δεδομένα είναι οργανωμένα, γεγονός που καθιστά ευκολότερη την ανάλυση τους σε σύγκριση με τα μη δομημένα δεδομένα. Το JSON και το XML είναι παραδείγματα ημι - δομημένων δεδομένων.

2.4.4. Ορθότητα (Veracity)

Η ορθότητα των δεδομένων αφορά την ορθότητα και την αξιοπιστία αυτών. Τα δεδομένα που θα χρησιμοποιηθούν είναι απαραίτητο να ελεγχθούν ως προς τα παραπάνω στοιχεία όσο τον δυνατόν καλύτερα, προκειμένου να διασφαλιστεί η ορθότητα των αποτελεσμάτων στα οποία θα μας οδηγήσουν.

Κάθε σύνολο δεδομένων εμπεριέχει έναν βαθμό θορύβου. Ο θόρυβος είναι μια πληροφορία χωρίς αξία και εφόσον αφαιρεθεί, απομένει μόνο η πληροφορία που μας ενδιαφέρει και μπορεί να αποδειχθεί ιδιαίτερα χρήσιμη. Αυτά τα δεδομένα που έχουν μεγαλύτερο όγκο πληροφορίας προς τον θόρυβο είναι και αυτά που θεωρούμε πιο έγκυρα. Ωστόσο σημαντικό ρόλο στην εγκυρότητα των δεδομένων έχει η προέλευσή τους αλλά και ο τρόπος με τον οποίο γίνεται η συλλογή τους (Shannon C., Weaver W., 1963).

2.4.5. Αξία (Value)

Η πρόσβαση στα Big Data αν δεν μπορεί να μετατραπεί σε αξία καταλήγει να είναι άχρηστη. Για αυτό και σε πολλές περιπτώσεις η αξία θεωρείται ως η σημαντικότερη από τα παραπάνω χαρακτηριστικά. Οι οργανισμοί καλούνται να επιλέξουν την πιο αποτελεσματική λύση, σύμφωνα πάντα από την πλευρά του κόστους, με στόχο την αξιοποίηση της πληροφορίας που θα οδηγήσει στην έγκαιρη και στην όσο το δυνατό πιο σωστή λήψη αποφάσεων, δίνοντας την μέγιστη δυνατή αξία στην επιχείρηση. Τέλος η αξία είναι και το χαρακτηριστικό που συνδέεται άμεσα με την ορθότητα των δεδομένων. Όσο μεγαλύτερη είναι η εγκυρότητα των δεδομένων τόσο μεγαλύτερη είναι και η αξία τους για την επιχείρηση.

Μέρος Β:

Η ανάλυση Big Data έχει μελετηθεί από μια σειρά επιστημονικών ερευνών, κριτικών, δημοσιεύσεων και ερευνητικών εργασιών. Μερικές από αυτές τις μελέτες εξετάζουν τις πρωταρχικές δυσκολίες και την τρέχουσα κατάσταση των Big Data. Για παράδειγμα, οι Chen et al. διεξήγαγαν μια ενδελεχή εξέταση τεχνολογιών που σχετίζονται με Big Data, συμπεριλαμβανομένου του Cloud Computing, του Internet of Things και των κέντρων δεδομένων. Όσον αφορά την αναπαράσταση και τη μείωση δεδομένων, τον κύκλο ζωής, τη διαχείριση ενέργειας και την επεκτασιμότητα, οι επιστήμονες συζήτησαν τις δυσκολίες των μεγάλων δεδομένων (Amadio & Haywood, 2019). Ειδικότερα επικεντρώθηκαν σε όλες τις πτυχές της αλυσίδας αξίας Big Data, συμπεριλαμβανομένης της δημιουργίας δεδομένων, της απόκτησης (συλλογή δεδομένων, μεταφορά και προεπεξεργασία), αποθήκευση σε καταναμημένα συστήματα ή βάσεις δεδομένων NoSQL και ανάλυση των δεδομένων (Bock, 2016).

Μία από τις πιο ευρέως χρησιμοποιούμενες τεχνολογίες λογισμικού στις μέρες μας είναι τα υπολογιστικά φύλλα. Έχουν αποδειχθεί ότι είναι χρήσιμα για την οργάνωση, το χειρισμό, την εξέταση και την οπτικοποίηση δεδομένων καθώς και για την εκτέλεση αριθμητικών υπολογισμών. Αλλά μέχρι τώρα, οι μελέτες πεδίου επικεντρώνονταν κυρίως σε ένα χαρακτηριστικό των υπολογιστικών φύλλων: τους τύπους. Εδώ, περιγράφουμε μια μελέτη πεδίου τριών μερών που διευρύνει αυτήν την εστίαση προκειμένου να εντοπίσει ένα ευρύτερο φάσμα δυσκολιών που αντιμετωπίζουν οι άνθρωποι κατά τη δημιουργία και τη χρήση υπολογιστικών φύλλων. Αυτή η μελέτη εντόπισε μια σειρά από ευκαιρίες για τη βελτίωση των επεξεργαστών υπολογιστικών φύλλων, όπως η δημιουργία νέων τρόπων δημιουργίας υπολογιστικών φύλλων, η ενίσχυση της υποστήριξης για επαναχρησιμοποίηση υπολογιστικών φύλλων και η διευκόλυνση της ανακάλυψης και χρήσης λειτουργιών για τους χρήστες (Chambers & Scaffidi, 2008).

Η διερεύνηση των συνδέσμων σε πλούσιες, περίπλοκες πηγές δεδομένων είναι ο στόχος της ποιοτικής έρευνας. Τα στατιστικά εργαλεία των ποσοτικών προσεγγίσεων διαχωρίζουν τα σημαντικά στοιχεία των δεδομένων με τρόπο αντιπαραγωγικό (Chen, Mao, & Liu, 2014). Ωστόσο, όπως και οι ποσοτικοί ερευνητές, οι ποιοτικοί ερευνητές συχνά εξακολουθούν να αντιμετωπίζουν δυσκολίες λόγω του πλήθους των δεδομένων και χρειάζονται τα απαραίτητα εργαλεία και τις κατάλληλες μεθόδους για να τα χειριστούν. Η ανάγκη αυτή οδήγησε στη δημιουργία ενός αριθμού προγραμμάτων λογισμικού που προορίζονταν για αυτή τη χρήση. Το Microsoft Excel, ωστόσο, είναι μια επιλογή που συχνά αγνοείται. Το Excel αναφέρεται συχνά ως «θρυμματιστής» αριθμών, ωστόσο, η ποιοτική ανάλυση μπορεί να πραγματοποιήσει χρήση της δομής, της επεξεργασίας δεδομένων και των χαρακτηριστικών εμφάνισης (Meyer & Avery, 2008).

Οι τύποι, ειδικά η ακρίβεια των τύπων, έχουν λάβει το μεγαλύτερο μέρος της προσοχής σε προηγούμενες επιτόπιες μελέτες σχετικά με τη χρήση υπολογιστικών φύλλων. Σύμφωνα με αυτήν την έρευνα, τα υπολογιστικά φύλλα συνήθως "παράγονται με ανεπίσημο, επαναληπτικό τρόπο" χωρίς ποιοτικό έλεγχο (Cragg & King, 1993), με έως και το 94% των υπολογιστικών φύλλων να έχουν συνήθως τουλάχιστον ένα

σφάλμα τύπου (Panko, 2005). Υπάρχουν πολλοί διαφορετικοί τύποι σφαλμάτων που μπορεί τελικά να οδηγήσουν σε σημαντικές οικονομικές απώλειες (Caulkins et al., 2007). Αυτά τα ευρήματα καταδεικνύουν ότι η ορθότητα των τύπων είναι ένας ουσιαστικός στόχος που αξίζει ενδελεχής έρευνας, αν και οι τύποι αποτελούν μόνο ένα μικρό μέρος των δυνατοτήτων που παρέχονται από τους συντάκτες υπολογιστικών φύλλων (Panko, 2005).

Η εκκαθάριση, η μετατροπή και η ανάλυση ακατέργαστων δεδομένων είναι η διαδικασία απόκτησης χρήσιμων, κατάλληλων πληροφοριών που μπορούν να βοηθήσουν τις επιχειρήσεις να λάβουν τεκμηριωμένες αποφάσεις. Η τεχνική χρησιμεύει στη μείωση των κινδύνων που εμπεριέχονται στη λήψη αποφάσεων παρέχοντας σχετικές πληροφορίες και γεγονότα, τα οποία εμφανίζονται συχνά ως γραφήματα, φωτογραφίες, πίνακες και γραφήματα (Agarwal, 2021).

Η ανάλυση δεδομένων περιλαμβάνει όχι μόνο την ανάλυση δεδομένων αλλά και τη συλλογή δεδομένων, την οργάνωση, την αποθήκευση και τα εργαλεία και τις τεχνικές που χρησιμοποιούνται για τη λεπτομερέστερη εξερεύνηση των δεδομένων καθώς και εκείνα που χρησιμοποιούνται για την επικοινωνία των αποτελεσμάτων, όπως εργαλεία οπτικοποίησης δεδομένων (Bock, 2016). Η ανάλυση δεδομένων, από την άλλη πλευρά, εστιάζει στη διαδικασία μετατροπής των ακατέργαστων δεδομένων σε χρήσιμα στατιστικά στοιχεία, πληροφορίες και επεξηγήσεις (Amadio & Haywood, 2019).

Τα ακόλουθα στάδια αποτελούν την επαναληπτική διαδικασία ανάλυσης δεδομένων (Chen, Mao, & Liu, 2014):

- Πρέπει να προσδιορίζονται οι απαιτήσεις δεδομένων
- Συλλογή δεδομένων
- Επεξεργασία δεδομένων
- Καθαρισμός Δεδομένων
- Ανάλυση Δεδομένων
- Μετάδοση δεδομένων

Στην εποχή των μεγάλων δεδομένων, οι Birch, Lyford-Smith και Guo (2018) εξέτασαν το μέλλον των υπολογιστικών φύλλων. Τα πιο δημοφιλή εργαλεία για την αποθήκευση, το χειρισμό και τη μοντελοποίηση δεδομένων είναι τα υπολογιστικά φύλλα. Έχουν εφαρμοστεί με επιτυχία σε κάθε πτυχή της εταιρικής διαδικασίας τα τελευταία 30 χρόνια χάρη στην παρουσία του (Chen, Mao, & Liu, 2014). Παραδόξως, τα υπολογιστικά φύλλα δεν έχουν εξελιχθεί τόσο πολύ τα τελευταία 30 χρόνια. Μια ποικιλία μεταβλητών μετατόπισης αρχίζουν να έχουν αντίκτυπο στην τεχνολογία υπολογιστικών φύλλων καθώς εισέρχεται στην τέταρτη δεκαετία της. Η ανάπτυξη και η εφαρμογή της τεχνολογίας υπολογιστικών φύλλων σίγουρα θα επηρεαστεί περαιτέρω από την άνοδο των μεγάλων δεδομένων, των υπολογιστών τελικού χρήστη και των

φορητών υπολογιστών (Bock, 2016). Σύμφωνα με έρευνα των Shigarov, Khristyuk και Mikhailov (2019), τα υπολογιστικά φύλλα χρησιμοποιούνται συχνά στην επιστήμη, τη μηχανική, τις επιχειρήσεις και άλλες δραστηριότητες.

Σε γενικές γραμμές, ακολουθούν ορισμένες συμβουλές ανάλυσης δεδομένων Excel (Ray, 2020) είναι οι εξής:

1. Ένα από τα πιο χρησιμοποιούμενα εργαλεία για την ανάλυση δεδομένων είναι το Microsoft Excel.
2. Η ανάλυση δεδομένων με χρήση του Excel παρέχει ως βάση για την επιστήμη δεδομένων με R ή Python.
3. Οι βασικές συναρτήσεις του Excel χρησιμοποιούνται για την εξέταση δεδομένων για επιχειρηματικές αναλύσεις.

Καταργήστε τις περιττές τιμές

Οι διπλότυπες τιμές σε έναν πίνακα μπορούν να εξαλειφθούν χρησιμοποιώντας ένα εργαλείο που συνοδεύει το Excel. Με βάση τις καθορισμένες στήλες, εξαλείφει τις διπλές τιμές από τον δεδομένο πίνακα. Για παράδειγμα, εάν επιλεγούν δύο στήλες, αναζητά διπλότυπες τιμές που περιλαμβάνουν τον ίδιο συνδυασμό δεδομένων και από τις δύο στήλες.

2.5. Ανάλυση Δεδομένων & Excel

Τα υπολογιστικά φύλλα είναι πλέον τόσο διαδεδομένα στην ανάλυση δεδομένων και στη λήψη αποφάσεων όσο και οι αριθμομηχανές. Οι αναλυτικές δεξιότητες και οι δεξιότητες μοντελοποίησης που βασίζονται στο Excel έχουν μεγαλύτερη ζήτηση από φοιτητές, επαγγελματίες και υπαλλήλους προσλήψεων καθώς περνάει κάθε χρόνο. Το Excel είναι μια απαίτηση για όποιον προσπαθεί να μοντελοποιήσει επιχειρηματικές προκλήσεις. Μπορούμε να συνδυάσουμε αυτές τις αναλύσεις σε μια ενιαία πλατφόρμα μοντελοποίησης χάρη σε υπολογιστικά φύλλα. Ως αποτέλεσμα, η προσπάθειά μας για μοντελοποίηση είναι:

1. Ανθεκτική - μια ισχυρή δομή που μπορεί να προβλέψει ποικίλη μελλοντική χρήση.
2. Ευέλικτο - ικανό να προσαρμοστεί καθώς το πρόβλημα αλλάζει και εξελίσσεται.
3. Κοινοποιήσιμα - μοντέλα που μπορούν να κοινοποιηθούν από διάφορα άτομα σε πολλά επίπεδα του οργανισμού, τα οποία συμμετέχουν όλοι στη διαδικασία επίλυσης προβλημάτων (Amadio & Haywood, 2019).

Επιπλέον, σε σύγκριση με άλλες πιο σύνθετες γλώσσες προγραμματισμού που χρησιμοποιούνται σε πολλά συστήματα μοντελοποίησης, ο τυπικός προγραμματισμός που είναι απαραίτητος για τα υπολογιστικά φύλλα είναι πιο εύκολος στην κατανόηση. Ωστόσο, το Excel περιέχει μια ενσωματωμένη, ισχυρή γλώσσα προγραμματισμού, που

ονομάζεται VBA, για την αντιμετώπιση της σποραδικής ανάγκης για πιο επίσημο προγραμματισμό (Visual Basic for Applications).

2.5.1. Πλεονεκτήματα Excel

Τα ακόλουθα είναι μερικά από τα κύρια πλεονεκτήματα που παρουσιάζει το Microsoft Excel στην ανάλυση των Μεγάλων Δεδομένων:

- Ιδανική μέθοδος αποθήκευσης δεδομένων

Το Excel είναι ο καλύτερος τρόπος για να αποθηκεύσει κανείς δεδομένα, ακόμα κι αν δεν χρησιμοποιήσει καμία από τις επιλογές ή τις λειτουργίες που παρέχει (Gurta, 2020). Εκατομμύρια άνθρωποι χρησιμοποιούν τα υπολογιστικά φύλλα της εφαρμογής που ονομάζεται Excel, έτσι, άλλοι μπορούν να έχουν γρήγορη πρόσβαση στα δεδομένα που αποθηκεύετε στο Excel και να μοιράζεστε μαζί τους (Bose, 2019).

- Είναι σε θέση να εκτελεί υπολογισμούς,

Το γεγονός ότι το Excel μπορεί να εκτελέσει υπολογισμούς το καθιστά το καλύτερο πρόγραμμα υπολογιστικών φύλλων. Κάθε χρήστης του Excel έχει πρόσβαση σε μια ολοκληρωμένη λίστα συναρτήσεων που μπορούν να χρησιμοποιηθούν για υπολογισμούς (Gurta, 2020). Μια συνάρτηση Excel είναι ένας προ-γραμμένος τύπος που, όταν εφαρμόζεται στις τιμές που εισάγονται, παράγει ένα συγκεκριμένο αποτέλεσμα. Ωστόσο, εκτός από αυτό, ο χρήστης μπορεί να σχεδιάσει τους δικούς του τύπους συγχωνεύοντας δύο ή περισσότερες συναρτήσεις μαζί ή πραγματοποιώντας απλούς υπολογισμούς με τελεστές (Data Analysis, 2018). *(γίνεται εκτενής αναφορά στην συνέχεια)*

- Εργαλεία ανάλυσης δεδομένων

Ο κύριος λόγος για τη συλλογή δεδομένων είναι η ανάλυση, αξιολόγηση και η ερμηνεία τους με στόχο την παροχή νέων επιστημονικών πληροφοριών. Μερικά από τα πιο αποτελεσματικά εργαλεία ανάλυσης δεδομένων είναι διαθέσιμα στο Excel (Meyer & Avery, 2008). Για παράδειγμα είναι εφικτό να εισάγει κανείς έναν συγκεντρωτικό πίνακα από μεγάλα δεδομένα ή δεδομένα με εκατοντάδες σειρές και να δημιουργήσει έναν συνοπτικό πίνακα. *(γίνεται εκτενής αναφορά στην συνέχεια)*

- Εύκολες απεικονίσεις δεδομένων με γραφήματα

Υπάρχουν πολλές επιλογές στο Excel για τη δημιουργία γραφημάτων και την οπτική παρουσίαση δεδομένων. Αυτοί οι συνήθεις τύποι γραφημάτων μπορούν εύκολα να εισαχθούν και να αποθηκευτούν εκ των προτέρων για χρήση στην παρουσίαση των δεδομένων (Bose, 2019). Ο χρήστης μπορεί να μελετήσει όλα τα γραφήματα που μπορούν να εισαχθούν μεταβαίνοντας στην καρτέλα Εισαγωγή και επιλέγοντας την ομάδα Γραφήματα. Επίσης με την χρήση της εντολής "προτεινόμενα γραφήματα" προτείνεται από το Microsoft Excel το καλύτερο γράφημα για χρήση βάση των δεδομένων του φύλλου εργασίας. Η επιλογή κατασκευής σύνθετων γραφημάτων χρησιμοποιώντας συγκεκριμένες τεχνικές είναι η τελευταία (Amadio & Haywood, 2019). *(γίνεται εκτενής αναφορά στην συνέχεια)*

- Εκτύπωση αναφορών

Οι αναφορές μπορεί να χρειαστεί να εκτυπωθούν για κοινή χρήση ή παρουσίαση, ανάλογα με τον χρήστη (Data Analysis, 2018). Η εκτύπωση με το Excel είναι απίστευτα προσαρμόσιμη, πιο συγκεκριμένα είναι εφικτή η αλλαγή του περιθωρίου, το μέγεθος του χαρτιού, ο προσανατολισμός και άλλα. Πριν από την εκτύπωση μιας αναφοράς, η σελίδα μπορεί να διαμορφωθεί χρησιμοποιώντας ένα ευρύ φάσμα ρυθμίσεων (Meyer & Avery, 2008).

- Δημιουργία πινάκων

Η ικανότητα του Excel να δημιουργεί πίνακες εργαλείων και πρότυπα είναι ένα από τα πιο δημοφιλή χαρακτηριστικά του (Gupta, 2020).

- Γλώσσα προγραμματισμού VBA

Η Visual Basic for Applications (VBA), μια γλώσσα κωδικοποίησης αποκλειστικά για το Microsoft Office, επιτρέπει την αυτοματοποίηση σχεδόν όλων των μη αυτόματων εργασιών του Excel.

- Καθαρισμός και Μετασχηματισμός Δεδομένων

Η αντιμετώπιση ακατέργαστων δεδομένων που πρέπει να καθαριστούν για να χρησιμοποιηθούν είναι σύνηθες φαινόμενο (Gupta, 2020). Η μετατροπή και ο καθαρισμός δεδομένων απαιτούν πολύ χρόνο και μπορεί να είναι μια επίπονη λειτουργία. Το Power query, ένα εργαλείο μετασχηματισμού και καθαρισμού δεδομένων που προσφέρεται από τη Microsoft ως λύση, επιτρέπει στους χρήστες να φορτώνουν δεδομένα από πολλές πηγές, να τα μετατρέπουν και στη συνέχεια να τα φορτώνουν ξανά στο φύλλο εργασίας (Meyer & Avery, 2008).

- Αποτελεσματική μοντελοποίηση και ανάλυση

Τα δεδομένα δεν είναι περιορισμένα, ωστόσο το Excel έχει όριο σειρών και στηλών. Η Microsoft ανέπτυξε το Power Pivot ως αποτέλεσμα, επιτρέποντας στους χρήστες να αποθηκεύουν τεράστια δεδομένα με εκατομμύρια σειρές και στη συνέχεια να εκτελούν υπολογισμούς εκεί (Bose, 2019).

2.6. Βασικές Λειτουργίες Ανάλυσης Δεδομένων του Excel

Σύμφωνα με την παραπάνω βιβλιογραφική ανασκόπηση στην ακόλουθη ενότητα παρουσιάζονται οι εντολές που χρησιμοποιούνται περισσότερο στην ανάλυση των δεδομένων μέσω του Excel και δίνονται παραδείγματα κατανόησης των ακόλουθων εντολών.

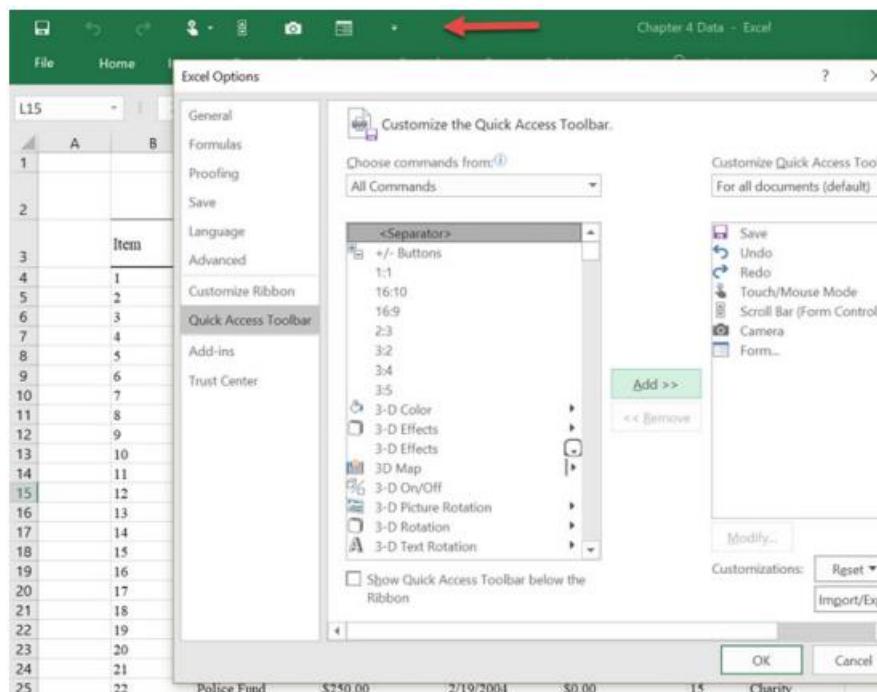
Το Excel είναι ένα ευέλικτο εργαλείο για την ανάλυση δεδομένων που μπορεί να χρησιμοποιηθεί για την εκτέλεση ενός ευρέος φάσματος εργασιών, από τον καθαρισμό και την προετοιμασία δεδομένων έως την προηγμένη στατιστική ανάλυση.

2.6.1. Data entry

Ξεκινάμε με τη διαδικασία απόκτησης δεδομένων, δηλαδή η διαδικασία λήψης δεδομένων από κάποια εξωτερική πηγή και τη μεταφορά της σε ένα φύλλο εργασίας του Excel. Η συγκεκριμένη διαδικασία είναι και από τις πρώτες που θα χρησιμοποιηθούν όταν πρόκειται να γίνει ανάλυση δεδομένων σε ένα υπολογιστικό φύλλο.

Το Excel έχει δύο πολύ χρήσιμα εργαλεία, Forms και Validation, που βοηθούν τον χρήστη να εισάγει δεδομένα με ακρίβεια. Η εισαγωγή δεδομένων είναι συχνά κουραστική και χωρίς ενδιαφέρον, και ως εκ τούτου, μπορεί εύκολα να οδηγήσει σε σφάλματα εισαγωγής. Αν εμείς πρόκειται να εισάγουμε έναν σχετικά μεγάλο όγκο δεδομένων, τότε αυτά τα εργαλεία μπορεί να είναι εξαιρετικά χρήσιμα. Μια εναλλακτική λύση είναι η εισαγωγή δεδομένων που μπορεί να έχουν αποθηκευτεί σε λογισμικό διαφορετικό από το Excel—μια βάση δεδομένων, ένα αρχείο κειμένου κ.λπ. Φυσικά, αυτό δεν αντικαθιστά την ανάγκη να εξεταστούν ενδελεχώς τα δεδομένα για σφάλματα.

Ας ξεκινήσουμε εξετάζοντας το εργαλείο Forms. Αυτό το εργαλείο επιτρέπει μια εξαιρετικά δομημένη και μέθοδος απόδειξης σφαλμάτων εισαγωγής δεδομένων. Το συγκεκριμένο εργαλείο δεν εμφανίζεται στο ribbon tab και μπορεί να προστεθεί στη γραμμή εργαλείων γρήγορης πρόσβασης που εμφανίζεται στο επάνω αριστερό μέρος του φύλλου εργασίας με την διαδικασία προσαρμογής της γραμμής εργαλείων γρήγορης πρόσβασης φαίνεται στην εικόνα που ακολουθεί:



Εικόνα 2: Excel Forms

Οι φόρμες στο Excel είναι μια δυνατότητα που επιτρέπει στους χρήστες να δημιουργούν προσαρμοσμένες έρευνες ή ερωτηματολόγια και να συλλέγουν απαντήσεις από άλλους χρήστες. Με τις Φόρμες, μπορείτε να δημιουργήσετε

διάφορους τύπους ερωτήσεων, όπως πολλαπλές επιλογές, βαθμολογία ή εισαγωγή κειμένου και μπορείτε να προσαρμόσετε την εμφάνιση της έρευνάς σας. Μόλις ολοκληρωθεί η φόρμα σας, μπορείτε να τη μοιραστείτε με άλλους για να συγκεντρώσετε απαντήσεις και τα αποτελέσματα συλλέγονται και αναλύονται αυτόματα σε ένα βιβλίο εργασίας του Excel.

Το Validations στο Excel είναι μια δυνατότητα που σας επιτρέπει να ελέγχετε τον τύπο και τη μορφή των δεδομένων που εισάγονται σε ένα κελί. Βοηθά να διασφαλιστεί ότι τα δεδομένα είναι ακριβή, συνεπή και συμμορφώνονται με συγκεκριμένα κριτήρια ή κανόνες. Υπάρχουν διάφοροι τύποι επικύρωσης δεδομένων που είναι διαθέσιμοι στο Excel, όπως:

- Επικύρωση ολόκληρου αριθμού: Αυτό διασφαλίζει ότι μόνο ακέραιοι αριθμοί εισάγονται στο κελί.
- Δεκαδική επικύρωση: Αυτό διασφαλίζει ότι μόνο δεκαδικοί αριθμοί εισάγονται στο κελί.
- Επικύρωση ημερομηνίας και ώρας: Αυτό διασφαλίζει ότι στο κελί εισάγονται μόνο έγκυρες ημερομηνίες και ώρες.
- Επικύρωση λίστας: Αυτό διασφαλίζει ότι μόνο συγκεκριμένες τιμές από μια προκαθορισμένη λίστα εισάγονται στο κελί.
- Προσαρμοσμένη επικύρωση: Αυτό σας επιτρέπει να δημιουργήσετε τους δικούς σας κανόνες επικύρωσης με βάση τις συγκεκριμένες ανάγκες σας.

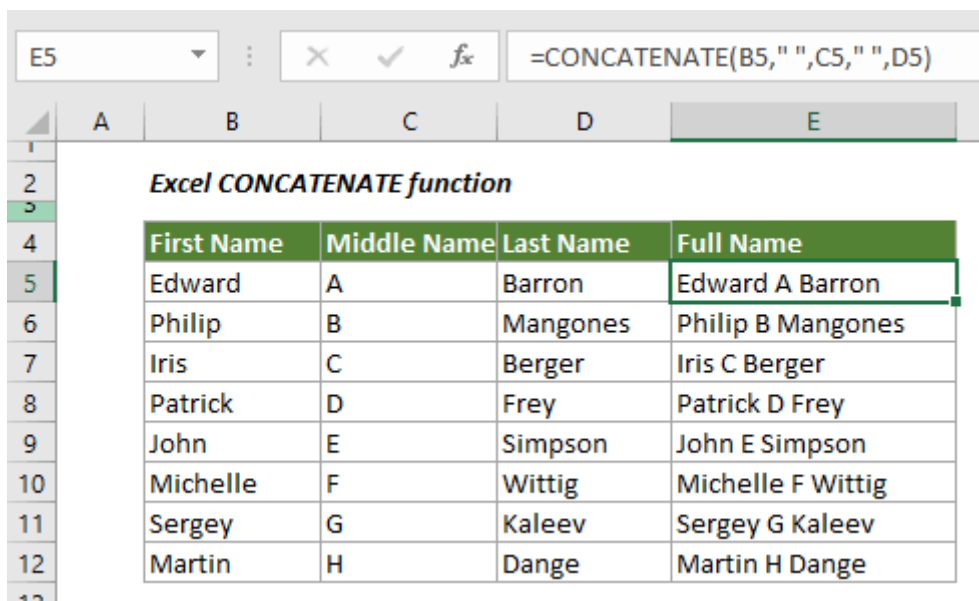
Για να ρυθμίσετε την επικύρωση δεδομένων στο Excel, ακολουθήστε τα εξής βήματα:

- Επιλέξτε το κελί ή το εύρος κελιών στα οποία θέλετε να εφαρμόσετε την επικύρωση.
- Μεταβείτε στην καρτέλα "Δεδομένα" στην κορδέλα και επιλέξτε "Επικύρωση δεδομένων".
- Στο παράθυρο διαλόγου "Επικύρωση δεδομένων", επιλέξτε τον τύπο επικύρωσης που θέλετε να εφαρμόσετε και εισαγάγετε τυχόν πρόσθετα κριτήρια ή κανόνες.
- Κάντε κλικ στο "OK" για να εφαρμόσετε την επικύρωση.

Μόλις εφαρμοστεί η επικύρωση, εάν ένας χρήστης προσπαθήσει να εισαγάγει δεδομένα που δεν πληρούν τα καθορισμένα κριτήρια, το Excel θα εμφανίσει ένα μήνυμα σφάλματος και θα αποτρέψει την εισαγωγή των δεδομένων. Η επικύρωση δεδομένων είναι ένα ισχυρό εργαλείο στο Excel που βοηθά στη διατήρηση της ακρίβειας και της συνέπειας των δεδομένων. Είναι ιδιαίτερα χρήσιμο όταν εργάζεστε με μεγάλα σύνολα δεδομένων ή όταν τα δεδομένα εισάγονται από πολλούς χρήστες. (Agarwal, 2021)

2.6.2. CONCATENATE

Ένας από τους πιο εύκολους στην κατανόηση αλλά και τους πιο αποτελεσματικούς τύπους για ανάλυση δεδομένων είναι η συνάρτηση Concatenate (Amadio & Haywood, 2019). Ένα μεμονωμένο κελί μπορεί να περιέχει κείμενο, αριθμούς, ημερομηνίες και άλλες πληροφορίες από πολλά κελιά τα οποία μπορούν να συνδυαστούν σε ένα. Αυτή είναι μια χρήσιμη μέθοδος για τη δημιουργία τελικών σημείων API, SKU προϊόντων και ερωτημάτων Java. Πιο συγκεκριμένα με αυτή τη συνάρτηση έχουμε την δυνατότητα να συνδέσουμε δύο ή περισσότερα στοιχεία κειμένου από πολλά κελιά σε ένα.



First Name	Middle Name	Last Name	Full Name
Edward	A	Barron	Edward A Barron
Philip	B	Mangones	Philip B Mangones
Iris	C	Berger	Iris C Berger
Patrick	D	Frey	Patrick D Frey
John	E	Simpson	John E Simpson
Michelle	F	Wittig	Michelle F Wittig
Sergey	G	Kaleev	Sergey G Kaleev
Martin	H	Dange	Martin H Dange

Εικόνα 3: Concatenate function Excel

2.6.3. LEN

Αυτή η συνάρτηση χρησιμοποιείται στην ανάλυση δεδομένων για την εμφάνιση του αριθμού των χαρακτήρων σε κάθε κελί (Meyer & Avery, 2008). Συνήθως χρησιμοποιείται κατά την προσπάθεια διάκρισης μεταξύ αριθμών ή κατά την εργασία με περιεχόμενο που έχει όριο χαρακτήρων. Επιστρέφει γρήγορα τον αριθμό των χαρακτήρων σε ένα δεδομένο κελί (Amadio & Haywood, 2019). Το ακόλουθο είναι ένα απλό παράδειγμα χρήσης την παραπάνω εντολής για την κατανόησή της:

String	Count	Formula	Description
4534	4	=LEN(C3)	Length of 4534
123.11	6	=LEN(C4)	Length of 123.11
23.00	2	=LEN(C5)	Length of 23(formatted as 23.00)
19-Aug-13	4	=LEN(C6)	Length of 4980(formatted as date)
Apple	5	=LEN(C7)	Length of Apple
Water Melon	11	=LEN(C8)	Length of Water Melon

Εικόνα 4: LEN Excel

2.6.4. NETWORKDAYS

Κατά την χρήση της συνάρτησης, επιστρέφεται ο αριθμός για τις πλήρεις εργάσιμες μέρες που μεσολαβούν ανάμεσα στις ημερομηνίες των ορισμάτων ημερομηνία_έναρξης και ημερομηνία_λήξης. Ο αριθμός των Σαββατοκύριακων εξαιρείται αυτόματα. Στο Excel, κατηγοριοποιείται ως συνάρτηση ημερομηνίας/ώρας (Aydin & Aydin, 2022). Στα χρηματοοικονομικά και τη λογιστική, η συνάρτηση καθαρής εργάσιμης ημέρας χρησιμοποιείται, μεταξύ άλλων, για τον υπολογισμό των παροχών των εργαζομένων με βάση τις ημέρες εργασίας, τον αριθμό των εργάσιμων ημερών που είναι προσβάσιμες σε ένα έργο ή τον αριθμό των εργάσιμων ημερών που απαιτούνται για την επίλυση ενός προβλήματος πελάτη (Amadio & Haywood, 2019). Η σύνταξη για την παραπάνω εντολή είναι η ακόλουθη:

=NETWORKDAYS (start_date, end_date, [holidays])

Start Date	End Date	Holidays
1/1/2022	31/12/2022	6/1/2022
		7/3/2022
		25/3/2022
		22/4/2022
		24/4/2022
		25/4/2022
		2/5/2022
		13/6/2022
		15/8/2022
		28/10/2022
Working Days	251	

Εικόνα 5: NETWORKDAYS Excel

Στο παράδειγμα αυτό μετρήσαμε όλες τις εργάσιμες μέρες για το 2022. Στην στήλη των Holidays είναι όλες οι αργίες του έτους που δεν τυχαίνουν να είναι Σάββατο ή Κυριακή. Με την συνάρτηση NETWORKDAYS, εφόσον αφαιρεί αυτόματα τα Σαββατοκύριακα, προστέθηκαν και οι αργίες τις οποίες διευκρινίσαμε και προέκυψε το τελικό νούμερο των 251 εργάσιμων ημερών.

2.6.5. COUNTIF / COUNTIFS

Αποτελεί ένα πολύ αποτελεσματικό εργαλείο για την ανάλυση δεδομένων του Excel αυτή η συνάρτηση. Η λειτουργία της είναι παρόμοια με αυτή της SUMIFS (Meyer & Avery, 2008). Η συνάρτηση χαρακτηρίζεται ως στατική και παρακολουθεί πόσες τιμές πληρούν ένα δεδομένο σύνολο κριτηρίων, επομένως, δεν απαιτεί ένα εύρος αθροίσματος όπως το SUMIFS. Η χρήση της μπορεί να είναι απλή αλλά και πιο εξειδικευμένη όπως στα ακόλουθα παραδείγματα:

Name	Count
Lily James	2
Lily Core	
Teddy Kristin	
Lily James	
Grace Brown	

Εικόνα 6: COUNTIFS Excel (1)

Name	Count
Lily James	3
Lily Core	
Teddy Kristin	
Lily James	
Grace Brown	

Εικόνα 7: COUNTIFS Excel (2)

Στο συγκεκριμένο παράδειγμα έχουμε μια απλή χρήση της συγκεκριμένης συνάρτησης όπου στην περίπτωση 1 αναγνωρίζει τον αριθμό των φορών εμφάνισης ενός ονοματεπωνύμου ενώ στην 2^η περίπτωση επιλέγουμε μόνο τον αριθμό φορών εμφάνισης ενός συγκεκριμένου ονόματος.

Name List	Criteria1	Criteria2	Count
James Band	James Band	Steven White	4
Coco Li			
James Band			
Steven White			
Vivien Kate			
Steven White			

Εικόνα 8: COUNTIFS Excel (3)

Στο παραπάνω παράδειγμα έχουμε μια πιο σύνθετη χρήση την συγκεκριμένης συνάρτησης. Εάν θέλουμε να μετρήσουμε τα κελιά τα οποία πληρούν τουλάχιστον μία από τις δεδομένες συνθήκες, δηλαδή να είναι ίσα με την τιμή x ή y μπορεί να γίνει την χρήση πολλαπλών COUNTIF αλλά και με την χρήση του SUM και έπειτα την συνάρτηση COUNTIF για να είναι πιο σύντομη η εντολή. Και στις δύο περιπτώσεις το αποτέλεσμα που προκύπτει είναι το ίδιο.

2.6.6. COUNTA

Η κατάσταση ενός κελιού ως κενού ή όχι καθορίζεται από το αυτή τη συνάρτηση (Agarwal, 2021). Σύμφωνα με την βιβλιογραφία ένας αναλυτής δεδομένων, συναντά συχνά ελλιπή σύνολα δεδομένων τα οποία μπορεί να διερευνήσει για τυχόν κενά στο σύνολο δεδομένων χρησιμοποιώντας αυτή την συνάρτηση χωρίς να χρειάζεται να αναδιατάξει τα δεδομένα. Πιο συγκεκριμένα μας δίνει την δυνατότητα καταμέτρησης κελιών που των κελιών που περιέχουν οποιοδήποτε τύπο πληροφοριών, συμπεριλαμβανομένων τιμών σφάλματος και κενού κειμένου (""). Η συγκεκριμένη συνάρτηση δεν καταμετρά τα κενά κελιά. (Aydin & Aydin, 2022).

Η χρήση της γίνεται σε πολλές περιπτώσεις σε συνδυασμό με την εντολή COUNT (σε περιπτώσεις που μας ενδιαφέρουν τα κελιά που περιέχουν μόνο αριθμούς) αλλά και με τις εντολές COUNTIF / COUNTIFS που αναφέραμε παραπάνω. Έτσι επιτυγχάνεται η αποκόμιση των δεδομένων που χρειάζεται ο χρήστης από ένα μεγάλο σύνολο δεδομένων.

2.6.7. SUMIF / SUMIFS

Το SUMIF είναι μια συνάρτηση του Excel που αθροίζει τις τιμές σε μια περιοχή που πληρούν ένα συγκεκριμένο κριτήριο. Η σύνταξη είναι =SUMIF(εύρος, κριτήριο, [εύρος_άθροισης]). Θα αθροίσει τις τιμές στα αντίστοιχα κελιά του εύρους sum_, ή του εύρους αν παραλειφθεί το sum_range, που πληρούν το κριτήριο.

Το SUMIFS είναι μια παρόμοια λειτουργία που σας επιτρέπει να καθορίσετε πολλά κριτήρια. Η σύνταξη είναι =SUMIFS(άθροισμα_εύρος, περιοχή 1, κριτήριο 1, περιοχή 2, κριτήριο 2, ...). Θα αθροίσει τις τιμές στο sum_range όπου τα αντίστοιχα κελιά στο εύρος 1 πληρούν το κριτήριο 1 και στο εύρος 2 το κριτήριο 2 και ούτω καθεξής.

Στην συνέχεια έχουμε παράδειγμα της παραπάνω εντολής να εκτελείται με πολλαπλά κριτήρια κάθε φορά για να γίνει πλήρως κατανοητή η λειτουργία της εντολής.

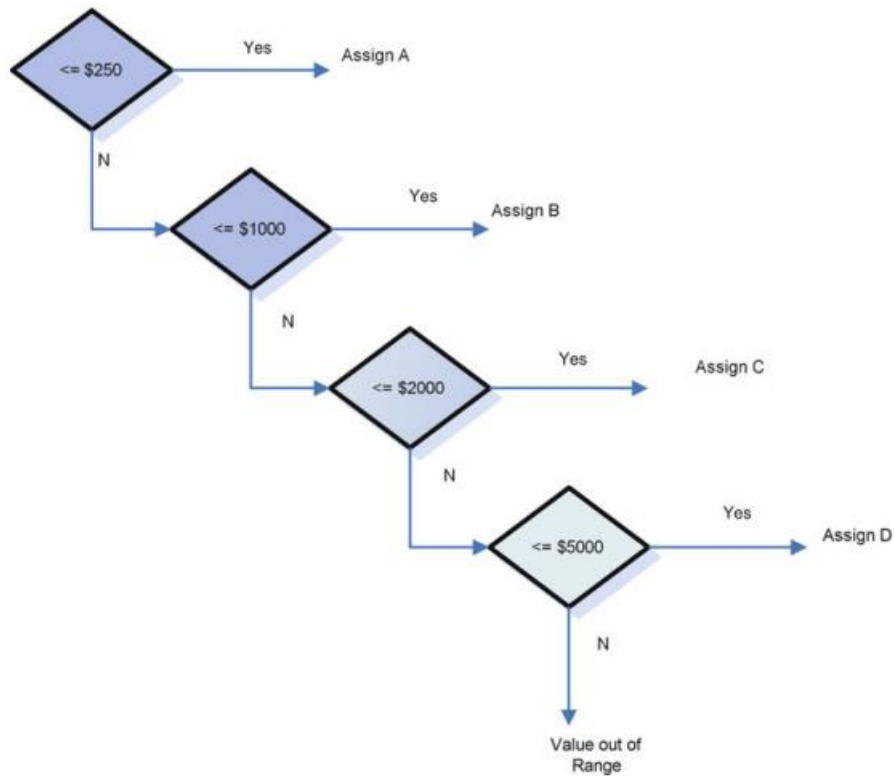
	A	B	C	D	E	F	G	H	I
1	Product	Sold Today	SUMIF to sum numbers in column B based on multiple column A						
2	Apples	31	Example 1 -- SUMIF in its simplest form, with 1 criteria:						
3	Oranges	33	Criteria	Apples					
4	Pears	25	Result	315					
5	Oranges	27	Formula	=SUMIF(A2:A23,E4,B2:B23)					
6	Bananas	14	Example 2 -- SUMIF for two criteria using SUMIF twice, clunky:						
7	Pineapples	71	Criteria	Apples	Oranges				
8	Apples	72	Result	554					
9	Oranges	78	Formula	=SUMIF(A2:A23,E9,B2:B23)+SUMIF(A2:A23,F9,B2:B23)					
10	Bananas	68	SUMIF with SUMPRODUCT for multiple hard-entered criteria.						
11	Pineapples	33	Example 3 -- SUMIF with SUMPRODUCT for 2 hard-entered criteria:						
12	Bananas	7	Criteria	Apples	Oranges				
13	Apples	72	Result	554					
14	Oranges	76	Formula	=SUMPRODUCT(SUMIF(A2:A23,{"Apples","Oranges"},B2:B23))					
15	Bananas	57	Example 4 -- SUMIF with SUMPRODUCT for 3 hard-entered criteria:						
16	Apples	69	Criteria	Apples	Oranges	Pears			
17	Oranges	25	Result	759					
18	Pears	42	Formula	=SUMPRODUCT(SUMIF(A2:A23,{"Apples","Oranges","Pears"},B2:B23))					
19	Pineapples	1							
20	Apples	71							
21	Pears	54							
22	Pears	84							
23	Bananas	74							

Εικόνα 9: SUMIF / SUMIFS Excel

2.6.8. IF

Η λειτουργία της IF είναι αρκετά χρήσιμη, καθώς με τη βοήθεια αυτής της δυνατότητας, ενδέχεται να λαμβάνονται αυτόματα αποφάσεις για τα υπολογιστικά φύλλα. Με βάση το αποτέλεσμα μιας λογικής δοκιμής, θα μπορεί έναν αναλυτή χρησιμοποιήσει το IF για να δώσει εντολή στο Excel να πραγματοποιήσει έναν νέο υπολογισμό ή να εμφανίσει μια διαφορετική τιμή (μια απόφαση). Η συνάρτηση IF ζητά από τον χειριστή να υλοποιήσει μια λογική δοκιμή και να καθορίσετε ως προς το τι πρέπει να πραγματοποιήσει εάν η δοκιμή είναι επιτυχής ή αν είναι ανεπιτυχής (Agarwal, 2021).

Στην συνέχεια δίνεται και διαγραμματικά ο τρόπος με τον οποίο λειτουργεί ακριβώς η εντολή αυτή για την ευκολότερη κατανόησή της:



Εικόνα 10:IF Excel

Το συγκεκριμένο παράδειγμα αναφέρεται σε συνδυασμό πολλαπλών IF που χρησιμοποιούνται ταυτόχρονα για να επιτευχθεί το επιθυμητό αποτέλεσμα για τον χρήστη.

2.6.9. VLOOKUP & HLOOKUP

Η συνάρτηση VLOOKUP χρησιμοποιείται για να αναζητήσει μια τιμή σε έναν πίνακα και να επιστρέψει μια αντίστοιχη τιμή. Στον παραπάνω πίνακα που περιλαμβάνει πολιτικές και πελάτες, η συνάρτηση VLOOKUP μπορεί να χρησιμοποιηθεί για να συσχετίσει το όνομα μιας πόλης με έναν πελάτη βάσει ενός κοινού αναγνωρισμένου.

Policy details					Customer details					
Customer id	Policy id	Product id	Premium (\$)	City	Customer id	Name	Gender	Age	City	Vintage (in months)
1	4010	4	2300		1	Joe	M	33	London	21
2	1001	1	1500		2	Jon	M	30	Nottingham	10
3	2030	2	5200		3	Catherine	F	28	Birmingham	3
4	1020	1	5600		4	Lucy	F	36	Manchester	5
5	1025	1	18400		5	Alex	M	30	London	3
6	4003	4	1500		6	Andy	M	35	London	21
7	1090	1	17600		7	Natasha	F	32	London	11
8	4123	4	3000		8	Prasad	M	47	Edinburgh	20
9	1111	1	9600		9	Rita	F	29	Nottingham	15
10	1123	1	15800		10	Chris	M	28	Oxford	0
11	1233	1	6300		11	Anthony	M	38	Bradford	21
12	3456	3	18400		12	Mark	M	36	Manchester	20

Εικόνα 11: VLOOKUP Excel

Για να εκτελέσει αυτήν την εργασία, μπορεί να χρησιμοποιηθεί η συνάρτηση VLOOKUP με την ακόλουθη σύνταξη: =VLOOKUP(Κλειδί για αναζήτηση, Πίνακας πηγής, Αριθμός στήλης πηγής, Ακριβής Αντιστοιχίας). Για παράδειγμα, μπορεί να γραφεί η συνάρτηση =VLOOKUP(B4, \$H\$4:\$L\$15, 5, 0) στο κελί F4 για να επιστραφεί το όνομα της πόλης για όλα τα αναγνωρισμένα πελάτες.

Επίσης, είναι σημαντικό να κλειδώσετε τον πίνακα του πίνακα χρησιμοποιώντας το σύμβολο "\$" για να αποφύγετε σφάλματα αναφοράς κατά την αντιγραφή του τύπου προς τα κάτω. (Ray, 2020).

Η HLOOKUP είναι μια συνάρτηση του Excel για αναζήτηση, η οποία ανακτά δεδομένα από μια συγκεκριμένη σειρά σε έναν πίνακα. Αναζητά μια τιμή στην πρώτη σειρά του πίνακα και επιστρέφει μια άλλη τιμή στην ίδια στήλη από μια γραμμή σύμφωνα με την δεδομένη συνθήκη κάθε φορά (Agarwal, 2021). Πιο συγκεκριμένα χρησιμοποιεί κανείς αυτή τη συνάρτηση για να ελέγξει έναν συγκεκριμένο αριθμό σειρών σε έναν πίνακα δεδομένων όταν οι τιμές σύγκρισης βρίσκονται σε μια σειρά στην κορυφή, εάν οι τιμές σύγκρισης βρίσκονται σε μια στήλη που βρίσκεται αριστερά από τα δεδομένα που αναζητάτε (Aydin & Aydin, 2022). Για να γίνει κατανοητή η χρησιμότητα της εντολής αυτής, δίνεται τα παρακάτω παράδειγμα:

	A	B	C	D	E	F	G	H	I
1	Sales	10000	20000	30000	40000				
2	Unit	5%	7%	10%	15%				
3									
4	Seller	Sales	Unit	Formula					
5	Jhonson	10000	5%	=HLOOKUP(B5,\$A\$1:\$E\$2,2)					
6	Lily	15666		Description					
7	Coco	21002		look up value 10000 in table A1:E2, then return value in row2					
8	Teddy	30000							
9	Boosh	10822							
10	Hadern	30010							
11	Wyden	20934							

Εικόνα 12: HLOOKUP Excel

Στην συγκεκριμένη περίπτωση είναι η πιο απλή μορφή της συγκεκριμένης συνάρτησης που αν χρησιμοποιηθεί στα κελιά C5:C11 θα επιστρέφει όσες τιμές μπορούν να αντιστοιχηθούν από την γραμμή στην οποία "κοιτάει" η HLOOKUP. Σε πιο προχωρημένη μορφή της, μπορεί τα δεδομένα να γράφονται σε ένα κρυφό φύλλο του excel και από εκεί να υπάρχει αντιστοιχία με έναν πίνακα όπου θα έχει συμπληρωθεί η σωστή μορφή της HLOOKUP και θα παίρνουν τα κελία τις αντίστοιχες τιμές από το κρυφό φύλλο.

2.6.10. PivotTables & PivotCharts

Ένας συγκεντρωτικός πίνακας στο Excel είναι ένα ισχυρό εργαλείο σύνοψης δεδομένων που σας επιτρέπει να εξάγετε πληροφορίες και τάσεις από ένα μεγάλο σύνολο δεδομένων. Σας δίνει τη δυνατότητα να χειρίζεστε και να αναλύετε μεγάλες ποσότητες δεδομένων με ευέλικτο, διαδραστικό και δυναμικό τρόπο. Ένας συγκεντρωτικός πίνακας συνοψίζει δεδομένα ομαδοποιώντας και συγκεντρώνοντας δεδομένα με βάση επιλεγμένες μεταβλητές, όπως προϊόν, περιοχή, ημερομηνία κ.λπ. Το αποτέλεσμα είναι ένας νέος πίνακας που εμφανίζει μια σύνοψη των δεδομένων, συχνά με υποσύνολα και γενικά σύνολα. Αυτός ο συνοπτικός πίνακας μπορεί εύκολα να φιλτραριστεί, να ταξινομηθεί και να αναπτυχθεί για καλύτερη κατανόηση των υποκείμενων δεδομένων. Για να δημιουργήσετε έναν συγκεντρωτικό πίνακα, πρέπει να έχετε ένα σύνολο δεδομένων σε μορφή πίνακα με τα δεδομένα που θέλετε να αναλύσετε.

Στο παράδειγμα που ακολουθεί έχουν παρθεί τα δεδομένα από μια ομάδα 29 ατόμων από το σύνολο των ατόμων την έρευνας, για την αξιολόγηση 4 προϊόντων (4 ιστοσελίδες) ανάλογα με την χρησιμότητά τους και την διαδραστικότητά τους με τον χρήστη.

Αρχικά, τα δεδομένα της έρευνάς μας είναι το Φύλο (άνδρας ή θηλυκό) και η αξιολόγηση για το Προϊόν 1 (καλό ή κακό) Έτσι, εμείς μπορούμε για παράδειγμα να προσδιορίσουμε όλα τα θηλυκά που επιλέγουν το καλό ως γνώμη για το σχεδιασμό της ιστοσελίδας Προϊόν 1 ή παρομοίως, όλα τα αρσενικά που επιλέγουν το κακό ως γνώμη για το Προϊόν 1. Ο πίνακας περιλαμβάνει και τους 29 ερωτηθέντες, να κατανέμονται σε τέσσερις κατηγορίες με τέτοιο τρόπο έτσι ώστε να μην υπάρχει διπλότητα.

Η εικόνα που ακολουθεί, δείχνει τη διασταύρωση του Φύλου και της προτίμησης για το Προϊόν 1.

Product 1			
Count of Case	Column Labels		
Row Labels	bad	good	Grand Total
F	7	7	14
M	4	11	15
Grand Total	11	18	29

Εικόνα13:PivotTables & PivotCharts Excel (1)

Η εικόνα του πίνακα περιλαμβάνει και τους 29 ερωτηθέντες, να κατανέμονται σε τέσσερις κατηγορίες με τέτοιο τρόπο έτσι ώστε να μην υπάρχει διπλότητα.

Ο πίνακας μπορεί να εμπλουτιστεί και με περισσότερες πληροφορίες, όπως είναι η περιοχή των συμμετεχόντων από την συλλογή των δεδομένων και να πάρει την ακόλουθη μορφή:

Product 1			
Count of Case	Column Labels		
Row Labels	bad	good	Grand Total
east			
F	1	2	3
M	2	4	6
north			
F	1	1	2
M	1	4	5
south			
F	3	2	5
M	1	1	2
west			
F	2	2	4
M		2	2
Grand Total	11	18	29

Εικόνα 14: PivotTables & PivotCharts Excel (2)

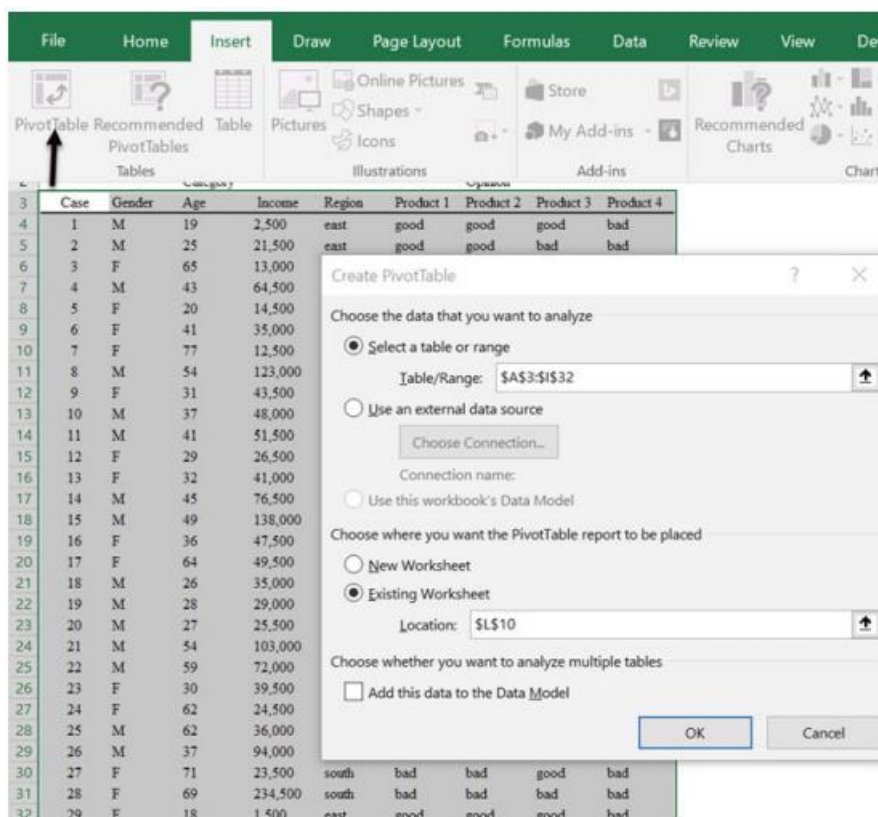
Στην συνέχεια αντί να δουλέψουμε με τον αριθμό των ερωτηθέντων, μπορούμε να το αντικαταστήσουμε με το μέσο εισόδημά τους και ανάλογα με την αξιολόγηση του προϊόντος 1 παρουσιάζοντας τα αποτελέσματα σε νέα εικόνα:

Product 1			
Average of Income	Column Labels		
Row Labels	bad	good	Grand Total
F	61,571.43	25,071.43	43,321.43
M	100,750.00	47,000.00	61,333.33
Grand Total	75,818.18	38,472.22	52,637.93

Εικόνα15: PivotTables & PivotCharts Excel (3)

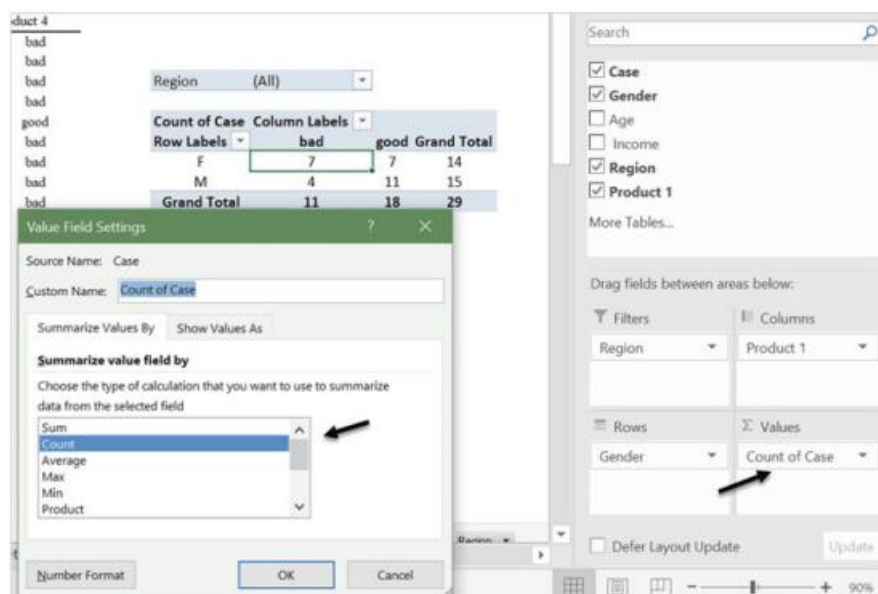
Μπορούμε επίσης να κατασκευάσουμε ένα παρόμοιο συγκεντρωτικό πίνακα για καθένα από τα τρία εναπομείναντα σχέδια ιστοσελίδων (Προϊόν 2, 3, 4). Οι συνδυασμοί που μπορούν να προκύψουν είναι πολλαπλοί σε κάθε έρευνα, ειδικότερα αν πρόκειται για μεγάλης έκτασης έρευνα. Προφανώς, αυτό θα μπορούσε να είναι μια κουραστική αγγαρεία, ειδικά εάν είναι ένα μεγάλο σύνολο δεδομένων, γι' αυτό εξαρτόμαστε από Συγκεντρωτικούς Πίνακες και τα Συγκεντρωτικά Διαγράμματα.

Αρχικά από την καρτέλα insert μπορούμε να προσθέσουμε τους συγκεντρωτικούς πίνακες και να ορίσουμε το εύρος των στοιχείων που θέλουμε να συμπεριλαμβάνει αλλά και το μέρος στο οποίο θα βλέπουμε τα αποτελέσματα αυτού:

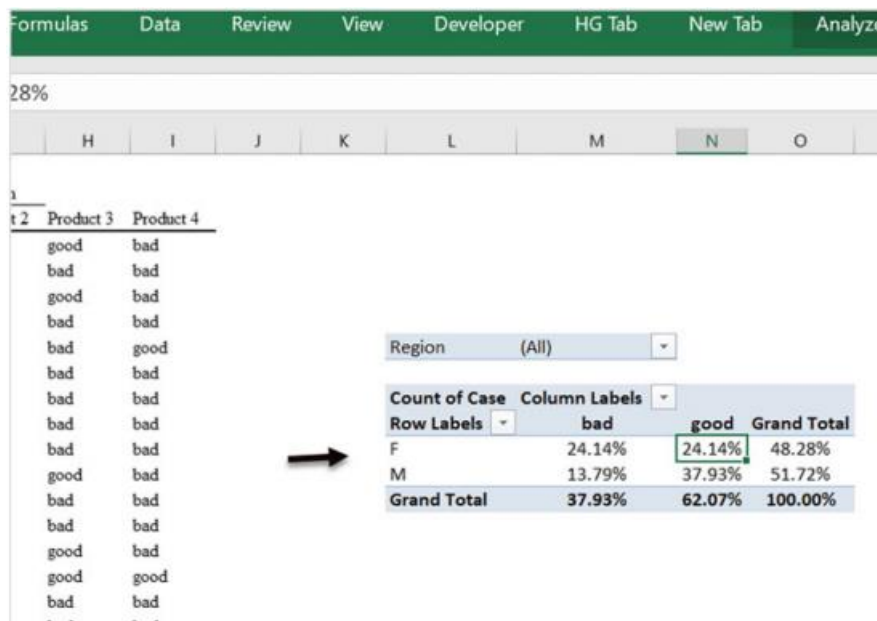


Εικόνα 16: PivotTables & PivotCharts Excel (4)

Στην συνέχεια οι επιλογές που μας δίνονται είναι πολλαπλές, ανάλογα τι αποτέλεσμα επιδιώκουμε να πετύχουμε. Αυτό μπορεί να σημαίνει απεικόνιση συγκεντρωτικών δεδομένων που μας ενδιαφέρουν όπως συμβαίνει στην εικόνα 17 αλλά και οι υπολογισμοί όπως μπορούμε να δούμε παρακάτω:

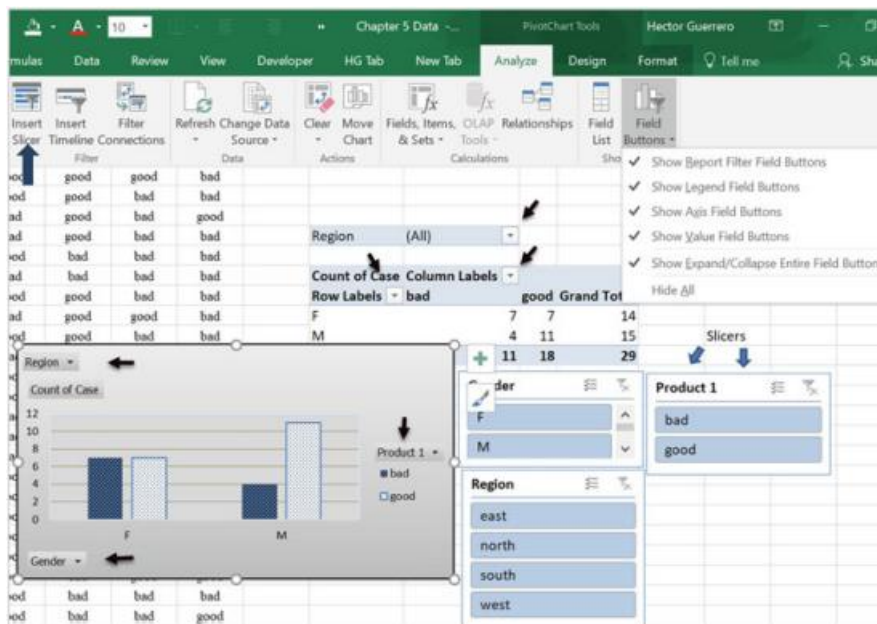


Εικόνα 17: PivotTables & PivotCharts Excel (5)

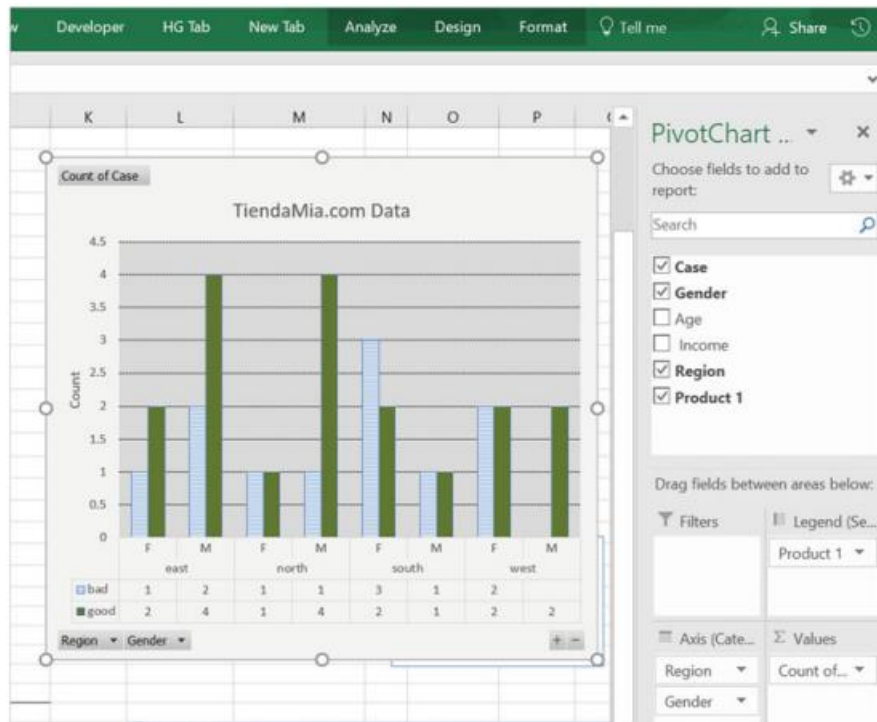


Εικόνα 18: PivotTables & PivotCharts Excel (6)

Πέρα από τους συγκεντρωτικούς πίνακες, μπορούμε να χρησιμοποιήσουμε και τα συγκεντρωτικά γραφήματα, τα οποία ακολουθούν την ίδια διαδικασία ακριβώς και μπορούν πολύ εύκολα να παράγουν συγκεντρωτικά γραφήματα για την οπτικοποίηση των αποτελεσμάτων μας, όπως συμβαίνει ακολούθως:



Εικόνα 19: PivotTables & PivotCharts Excel (7)



Εικόνα 20: PivotTables & PivotCharts Excel (8)

Σκοπός σύμφωνα με τα παραπάνω είναι να απαντηθούν οι ερωτήσεις:

- Εάν κάποια από τις ιστοσελίδες δείχνει να έχει την προτίμηση του κοινού, και
- Πως τα δημογραφικά δεδομένα σχετίζονται με τις προτιμήσεις των ερωτηθέντων.

Πριν ξεκινήσει η ανάλυση πρέπει να προειδοποιήσουμε πως το παραπάνω παράδειγμα έχει αρκετά μικρό δείγμα και είναι επικίνδυνο να συμπεράνουμε ότι το αποτέλεσμα ενός δείγματος είναι αντιπροσωπευτικό για την έρευνα εξ' ολοκλήρου. Σίγουρα η μελέτη σε μεγαλύτερο δείγμα παρέχει άνεση στην γενίκευση των αποτελεσμάτων και η συγκεκριμένη ανάλυση αποτελεί μια προκαταρκτική μελέτη.

Για την απάντηση της πρώτης ερώτησης δεν χρειάζεται να χρησιμοποιήσουμε τους συγκεντρωτικούς πίνακες. Ο συγκεντρωτικός πίνακας παρέχει μια εικόνα για το πώς τα διάφορα χαρακτηριστικά των ερωτηθέντων σχετίζονται με προτιμήσεις. Η ερώτησή μας αφορά συνοπτικά δεδομένα χωρίς να λαμβάνονται υπόψη τα λεπτομερή χαρακτηριστικά. Στην συγκεκριμένη περίπτωση θα χρειαστούμε την εντολή COUNTIF, όπου έχουμε αναλύσει παραπάνω, χρησιμοποιώντας τα κριτήρια που βλέπουμε στο ακόλουθο γράφημα:

Category						Opinion			
Case	Gender	Age	Income	Region	Product 1	Product 2	Product 3	Product 4	
1	M	19	2,500	east	good	good	good	bad	
2	M	25	21,500	east	good	good	bad	bad	
3	F	65	13,000	west	good	good	good	bad	
4	M	43	64,500	north	good	good	bad	bad	
5	F	20	14,500	east	bad	good	bad	good	
6	F	41	35,000	north	bad	good	bad	bad	
7	F	77	12,500	south	good	bad	bad	bad	
8	M	54	123,000	south	bad	bad	bad	bad	
21	M	54	103,000	south	good	bad	good	good	
22	M	59	72,000	west	good	good	good	bad	
23	F	30	39,500	west	good	bad	good	good	
24	F	62	24,500	east	good	bad	bad	bad	
25	M	62	36,000	east	good	bad	bad	good	
26	M	37	94,000	north	bad	bad	bad	bad	
27	F	71	23,500	south	bad	bad	good	bad	
28	F	69	234,500	south	bad	bad	bad	bad	
29	F	18	1,500	east	good	good	good	bad	
# of Bad=						11	14	16	24
# of Good=						18	15	13	5
Total=						29	29	29	29

Εικόνα 21: PivotTables & PivotCharts Excel (9)

Σύμφωνα λοιπόν με αυτή την ανάλυση καταλήγουμε εύκολα στο ότι υπάρχει η προτίμηση του κοινού για τον Προϊόν 1 χωρίς μεγάλη διαφορά από τα προϊόντα 2 και 3. Ωστόσο το 4^ο προϊόν είναι σίγουρα αρκετά μακριά στη σύγκριση με τα υπόλοιπα.

Η δεύτερη ερώτηση είναι ακριβώς ο τύπος της ερώτησης που μπορεί εύκολα να αντιμετωπιστεί με τη χρήση της ανάλυσης διασταυρούμενης πίνακα. Τα δημογραφικά μας χαρακτηριστικά αντιπροσωπεύονται από το Φύλο, Ηλικία, εισόδημα και περιοχή. Για το φύλο και την περιοχή έχουμε 2 και 4 πιθανές επιλογές αντίστοιχα. Για το εισόδημα και την ηλικία ωστόσο δεν υπάρχουν παραπάνω κατηγοριοποιήσεις. Εδώ προκύπτει το ερώτημα εάν είναι απολύτως απαραίτητο να έχουμε τέτοια λεπτομέρεια για τον στόχο μας να αναλύσουμε τη σύνδεση μεταξύ αυτών των δημογραφικών χαρακτηριστικών και τις προτιμήσεις για ιστοσελίδες. Και η απάντηση για να καταλήξουμε σε έγκυρα αποτελέσματα, είναι απολύτως απαραίτητη η ομαδοποίηση των δεδομένων αυτών.

Θα δημιουργήσουμε 2 κατηγορίες για την κάθε μεταβλητή, 18-37 ετών (τιμή 1) και 38 και άνω (τιμή 2), ενώ για το εισόδημα 0-37.000 δολάρια (τιμή 1) και 38.000 και άνω (τιμή 2). Υποθέτουμε πως αντιπροσωπεύουν επαρκώς τις ομάδες των καταναλωτών αυτές οι κατηγοριοποιήσεις, και με αυτό τον τρόπο προκύπτει ο παρακάτω πίνακας:

Case	Gender	Age	Age Cat	Income	Income Cat	Region	Product 1	Product 2	Product 3	Product 4
1	M	19	1	2,500	1	east	good	good	good	bad
2	M	25	1	21,500	1	east	good	good	bad	bad
3	F	65	2	13,000	1	west	good	good	good	bad
4	M	43	2	64,500	2	north	good	good	bad	bad
5	F	20	1	14,500	1	east	bad	good	bad	good
6	F	41	2	35,000	1	north	bad	good	bad	bad
7	F	77	2	12,500	1	south	good	bad	bad	bad
8	M	54	2	123,000	2	south	bad	bad	bad	bad
9	F	31	1	43,500	2	south	good	good	bad	bad
10	M	37	1	48,000	2	east	bad	good	good	bad
11	M	41	2	51,500	2	west	good	good	bad	bad
12	F	29	1	26,500	1	west	bad	good	bad	bad

Εικόνα 22: PivotTables & PivotCharts Excel (10)

Στην συνέχεια κατασκευάζονται οι 4 συγκεντρωτικοί πίνακες, ένας για κάθε προϊόν, και παρουσιάζονται ακολούθως δίνοντας έμφαση στις 2 κατηγοριοποιήσεις που προηγήθηκαν (φύλο και εισόδημα):

Count of Case		Product 1		Count of Case		Product 2	
Age Cat	Income Cat	bad	good	Age Cat	Income Cat	bad	good
1	1	2	6	1	1	1	7
	2	3	3		2	4	2
2	1	2	4	2	1	4	2
	2	4	5		2	5	4
Grand Total		11	18	Grand Total		14	15

Count of Case		Product 3		Count of Case		Product 4	
Age Cat	Income Cat	bad	good	Age Cat	Income Cat	bad	good
1	1	3	5	1	1	7	1
	2	3	3		2	5	1
2	1	4	2	2	1	5	1
	2	6	3		2	7	2
Grand Total		16	13	Grand Total		24	5

Εικόνα 23: PivotTables & PivotCharts Excel (11)

Σύμφωνα με τα αποτελέσματα αυτά μπορούμε να καταλήξουμε (όπως και παραπάνω), πως το προϊόν 4, ανεξαρτήτως ηλικίας και εισοδημάτων δεν έχει την προτίμηση του κοινού. Για τα υπόλοιπα προϊόντα τα συμπεράσματα είναι ασαφή δεδομένου ότι και το δείγμα είναι αρκετά μικρό. Αυτό μπορεί να αποκαλύψει νέα ερωτήματα τα οποία ίσως μας δώσουν την τελική απάντηση. Για να καταλήξουμε στην απάντηση πρέπει να κατηγοριοποιήσουμε τις απαντήσεις good/bad σε περισσότερες έτσι ώστε τα αποτελέσματα να είναι αντιπροσωπευτικά και λεπτομερή. Αυτό γίνεται παρακάτω με τις κατηγορίες 1/1 και 1/2 να αποτελούν την θετική προτίμηση και τις κατηγορίες 2/1 και 2/2 να αποτελούν την αρνητική προτίμηση.

Product	1\1	1\2	2\1	2\2	AVG	Wt. AVG
1	0.7500	0.5000	0.6667	0.5556	0.6181	0.6207
2	0.8750	0.3333	0.3333	0.4444	0.4965	0.5172
3	0.6250	0.5000	0.3333	0.3333	0.4479	0.4483
4	0.1250	0.1667	0.1667	0.2222	0.1701	0.1724
	1\1	1\2	2\1	2\2		
# Respondents in Categories	8	6	6	9	29	

Callout 1 (HG): $=(0.7500*8+0.5000*6+0.6667*6+0.5556*9)/\$F\$30$

Callout 2 (HG): $=AVERAGE(B25:E25)$

Εικόνα 24: PivotTables & PivotCharts Excel (12)

Σύμφωνα με την παραπάνω εικόνα, μπορούμε και επιβεβαιώνουμε (όπως και στην πρώτη ερώτηση) πως το προϊόν 1 είναι αυτό που έχει την προτίμηση του κόσμου, με το προϊόν 4 να μαζεύει τα περισσότερα αρνητικά σχόλια από το δείγμα μας.

2.7. Μέθοδοι – Τεχνικές ανάλυσης δεδομένων

Έπειτα από τις παραπάνω εντολές που αποτελούν σημαντικά μέρη της ανάλυσης των Big Data, υπάρχουν και οι ανάλογες τεχνικές που χρειάζονται μέχρι να επιτευχθεί το επιθυμητό αποτέλεσμα. Στην συνέχεια αναλύονται οι σημαντικότερες από αυτές.

2.7.1. Σειρές και Πίνακες και υποσύνολα με εύρη

Τα δεδομένα μπορεί να παρέχονται σε σχήμα πίνακα ή εύρους. Ορισμένες λειτουργίες μπορούν να πραγματοποιηθούν στα δεδομένα, ανεξάρτητα από το αν βρίσκονται σε πίνακα ή εύρος (Carey & Berk, 2007). Ωστόσο, ορισμένες τεχνικές λειτουργούν καλύτερα με δεδομένα που διατηρούνται σε πίνακες σε αντίθεση με τα εύρη. Επίσης μερικές λειτουργίες μπορούν να εκτελεστούν μόνο σε πίνακες. Έτσι, ένας αναλυτής οφείλει να αξιολογεί δεδομένα σε πίνακες και εύρη ανάλογα με την συνθήκη (Agarwal, 2021).

Σύμφωνα με την βιβλιογραφία, οι συγκεντρωτικοί πίνακες χρησιμοποιούνται συχνά για τη σύνοψη δεδομένων (Carey & Berk, 2007). Ωστόσο, τα υποσύνολα με εύρη είναι μια άλλη συνάρτηση του Excel που διευκολύνει την ομαδοποίηση και την κατάργηση ομαδοποίησης δεδομένων καθώς και τη σύνοψη δεδομένων σε εύρη.

2.7.2. Καθαρισμός Δεδομένων

Ο καθαρισμός δεδομένων, είναι η διαδικασία εντοπισμού και διόρθωσης ή αφαίρεσης σφαλμάτων, ασυνεπειών και ανακρίβειών σε ένα σύνολο δεδομένων. Είναι ένα σημαντικό βήμα στη διαδικασία προετοιμασίας δεδομένων, καθώς διασφαλίζει ότι τα δεδομένα είναι ακριβή, πλήρη και συνεπή, και επομένως κατάλληλα για ανάλυση ή άλλες χρήσεις.

Ο καθαρισμός δεδομένων μπορεί να περιλαμβάνει μια ποικιλία εργασιών, όπως αφαίρεση διπλότυπων, διόρθωση σφαλμάτων μορφοποίησης, συμπλήρωση τιμών που λείπουν, χειρισμός ακραίων τιμών ή ακραίων τιμών και τυποποίηση ή μετατροπή δεδομένων σε κοινή μορφή. Ο στόχος του καθαρισμού δεδομένων είναι να διασφαλίσει ότι τα δεδομένα είναι όσο το δυνατόν ακριβέστερα και αξιόπιστα, γεγονός που με τη σειρά του βοηθά στη βελτίωση της ποιότητας οποιωνδήποτε πληροφοριών, αποφάσεων ή αποτελεσμάτων που μπορεί να βασίζονται σε αυτά τα δεδομένα. (Carey & Berk, 2007)

Το Excel μπορεί να χρησιμοποιηθεί για τον καθαρισμό δεδομένων χρησιμοποιώντας τις διάφορες ενσωματωμένες λειτουργίες και δυνατότητες του. Ακολουθούν ορισμένα κοινά βήματα για τον καθαρισμό δεδομένων στο Excel:

- Κατάργηση διπλότυπων: Το Excel διαθέτει ένα ενσωματωμένο εργαλείο που μπορεί να αφαιρέσει διπλότυπες τιμές σε ένα σύνολο δεδομένων.
- Συμπλήρωση τιμών που λείπουν: Μπορείτε να χρησιμοποιήσετε τη συνάρτηση "γεμίματος" του Excel για να συμπληρώσετε γρήγορα τις τιμές που λείπουν σε ένα σύνολο δεδομένων, είτε αντιγράφοντας την τιμή από το παραπάνω κελί είτε χρησιμοποιώντας έναν προσαρμοσμένο τύπο.
- Διορθώστε τη μορφοποίηση: Το Excel μπορεί επίσης να βοηθήσει στη διόρθωση ζητημάτων μορφοποίησης, όπως ασυνεπής χρήση κεφαλαίων, κενά προπορευόμενα ή τελικά ή μορφές ημερομηνίας.
- Διαχωρισμός ή συγχώνευση κελιών: Εάν τα δεδομένα σε μια στήλη συνδυάζονται ή χωρίζονται σε πολλά κελιά, μπορείτε να χρησιμοποιήσετε τη συνάρτηση "διαίρεση" ή "συγχώνευση" του Excel για να αναδιοργανώσετε τα δεδομένα.
- Χρήση μορφοποίησης υπό όρους: Η μορφοποίηση υπό όρους του Excel μπορεί να επισημάνει ή να κωδικοποιεί κελιά με βάση συγκεκριμένα κριτήρια, διευκολύνοντας τον εντοπισμό ακραίων τιμών ή σφαλμάτων σε ένα σύνολο δεδομένων.

Αυτά είναι μόνο μερικά παραδείγματα των πολλών τρόπων που μπορεί να χρησιμοποιηθεί το Excel για τον καθαρισμό δεδομένων. Η συγκεκριμένη προσέγγιση θα εξαρτηθεί από τη φύση των δεδομένων και τις απαιτούμενες εργασίες καθαρισμού.

2.7.3. Μορφοποίηση υπό όρους

Η μορφοποίηση υπό όρους του Excel είναι μια δυνατότητα που επιτρέπει στους χρήστες να μορφοποιούν κελιά σε ένα υπολογιστικό φύλλο με βάση συγκεκριμένες συνθήκες ή κανόνες. Αυτό μπορεί να χρησιμοποιηθεί για την επισήμανση κελιών που πληρούν ορισμένα κριτήρια, όπως κελιά που περιέχουν μια συγκεκριμένη τιμή ή κελιά που βρίσκονται πάνω ή κάτω από ένα συγκεκριμένο όριο. Η μορφοποίηση υπό όρους μπορεί να βοηθήσει στην ευκολότερη ανάγνωση και ερμηνεία των δεδομένων και μπορεί να είναι ένα ισχυρό εργαλείο για την ανάλυση και την οπτικοποίηση δεδομένων.

Για να εφαρμόσετε μορφοποίηση υπό όρους στο Excel:

1. Επιλέξτε το εύρος των κελιών που θέλετε να μορφοποιήσετε.
2. Κάντε κλικ στην επιλογή "Μορφοποίηση υπό όρους" στην καρτέλα "Αρχική σελίδα" της κορδέλας.
3. Επιλέξτε τον τύπο μορφοποίησης που θέλετε να εφαρμόσετε, όπως κλίμακες χρώματος, γραμμές δεδομένων ή σύνολα εικονιδίων.
4. Ορίστε τα κριτήρια ή τους κανόνες για το πότε πρέπει να εφαρμόζεται η μορφοποίηση.
5. Κάντε κλικ στο "OK" για να εφαρμόσετε τη μορφοποίηση υπό όρους στα επιλεγμένα κελιά σας.

Μπορείτε επίσης να προσαρμόσετε τους δικούς σας κανόνες για μορφοποίηση υπό όρους επιλέγοντας "Νέος κανόνας" στο μενού Μορφοποίηση υπό όρους.

2.7.4. Ταξινόμηση και φιλτράρισμα

Για να προετοιμαστεί η ανάλυση δεδομένων ή για να εμφανιστεί ένα συγκεκριμένο κομμάτι κρίσιμων δεδομένων, ίσως χρειαστεί πρωταρχικά να ταξινομηθούν ή/και να φιλτραριστούν τα δεδομένα. Οι απλές δυνατότητες ταξινόμησης και φιλτραρίσματος του Excel μπορούν να χρησιμοποιηθούν για την εκτέλεση της ίδιας εργασίας (Agarwal, 2021). Οι πιο συχνά χρησιμοποιούμενες συναρτήσεις του Excel είναι η ταξινόμηση και το φιλτράρισμα. Ειδικότερα μπορεί κανείς να επιλέξει να υλοποιήσει ταξινόμηση των στηλών με αύξουσα ή φθίνουσα σειρά. Οι λίστες μπορούν να παραχθούν τυχαία, να ταξινομηθούν ανά χρώμα ή και τα δύο. Για την παρουσίαση δεδομένων που συμμορφώνονται με τα πρότυπα, χρησιμοποιούνται φίλτρα. Ορισμένες από τις επιλογές περιλαμβάνουν κατάργηση διπλότυπων, φόρμα δεδομένων, φίλτρα ημερομηνίας, φίλτρα αριθμού και κειμένου, περιγραφή δεδομένων (Amadio & Haywood, 2019).

Πιο συγκεκριμένα, το φιλτράρισμα στο Excel είναι μια δυνατότητα που σας επιτρέπει να εμφανίζετε μόνο ένα συγκεκριμένο σύνολο δεδομένων σε έναν πίνακα ή ένα εύρος με βάση ορισμένα κριτήρια. Αυτό μπορεί να είναι χρήσιμο για την ανάλυση και την εργασία με μεγάλα σύνολα δεδομένων. Υπάρχουν δύο τύποι φιλτραρίσματος στο Excel

- κανονικό φιλτράρισμα και προηγμένο φιλτράρισμα. Το κανονικό φιλτράρισμα, γνωστό και ως αυτόματο φιλτράρισμα, είναι η απλούστερη από τις δύο επιλογές. Για να εφαρμόσετε ένα κανονικό φίλτρο στο Excel, ακολουθήστε τα εξής βήματα:

- Επιλέξτε τον πίνακα ή το εύρος δεδομένων που θέλετε να φιλτράρετε.
- Μεταβείτε στην καρτέλα "Δεδομένα" στην κορδέλα και κάντε κλικ στο κουμπί "Φίλτρο".
- Κάντε κλικ στο αναπτυσσόμενο βέλος στη γραμμή κεφαλίδας της στήλης που θέλετε να φιλτράρετε και επιλέξτε τα κριτήρια με τα οποία θέλετε να φιλτράρετε. Μπορείτε να φιλτράρετε κατά κείμενο, αριθμούς, ημερομηνίες ή συγκεκριμένες τιμές.
- Επαναλάβετε το βήμα 3 για τυχόν πρόσθετες στήλες που θέλετε να φιλτράρετε.

Για να διαγράψετε τα φίλτρα, κάντε ξανά κλικ στο κουμπί "Φίλτρο" και επιλέξτε "Διαγραφή φίλτρου".

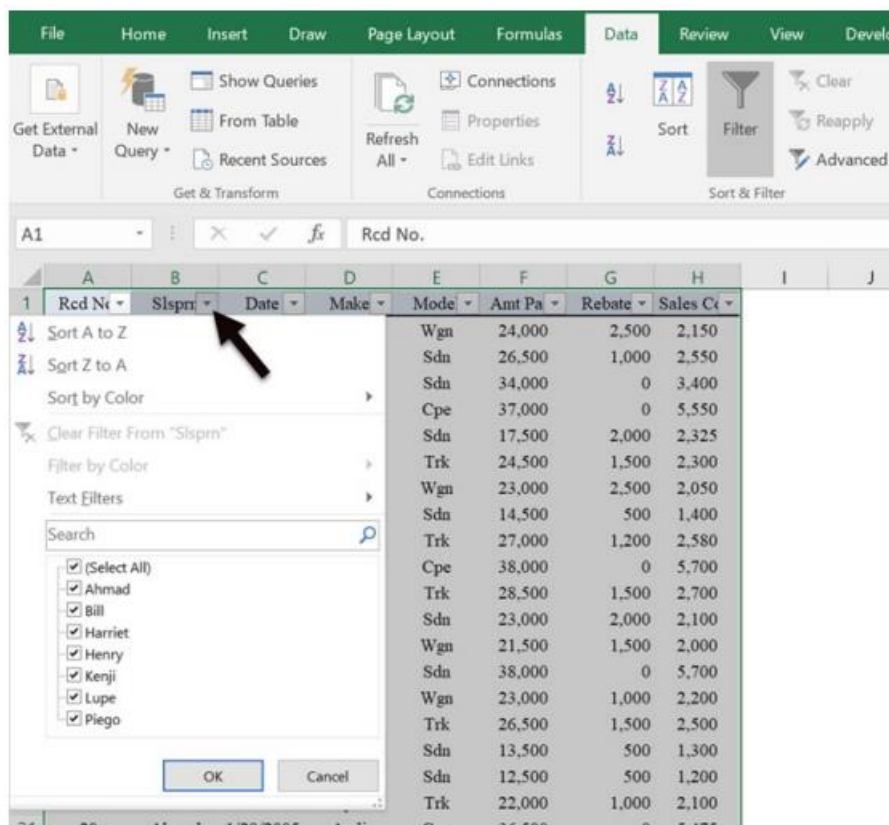
Το προηγμένο φιλτράρισμα σας επιτρέπει να δημιουργείτε πιο περίπλοκα φίλτρα χρησιμοποιώντας κριτήρια που εκτείνονται σε πολλές στήλες. Για να εφαρμόσετε ένα προηγμένο φίλτρο στο Excel, ακολουθήστε τα εξής βήματα:

- Δημιουργήστε ένα εύρος κριτηρίων ρυθμίζοντας μια ξεχωριστή περιοχή στο φύλλο εργασίας σας με τα κριτήρια που θέλετε να χρησιμοποιήσετε για το φίλτρο σας.
- Επιλέξτε τον πίνακα ή το εύρος δεδομένων που θέλετε να φιλτράρετε.
- Μεταβείτε στην καρτέλα "Δεδομένα" στην κορδέλα και κάντε κλικ στο κουμπί "Για προχωρημένους".
- Στο πλαίσιο διαλόγου "Σύνθετο φίλτρο", επιλέξτε το εύρος κριτηρίων και το εύρος των δεδομένων που θέλετε να φιλτράρετε.
- Επιλέξτε το πλαίσιο "Μόνο μοναδικές εγγραφές" εάν θέλετε να καταργήσετε τυχόν διπλότυπες τιμές.

Κάντε κλικ στο "OK" για να εφαρμόσετε το σύνθετο φίλτρο.

Το προηγμένο φιλτράρισμα μπορεί να είναι πιο περίπλοκο από το κανονικό φιλτράρισμα, αλλά προσφέρει περισσότερο έλεγχο στα δεδομένα που εμφανίζονται. Είναι χρήσιμο για εργασία με μεγάλα σύνολα δεδομένων που απαιτούν πιο συγκεκριμένα κριτήρια φιλτραρίσματος.

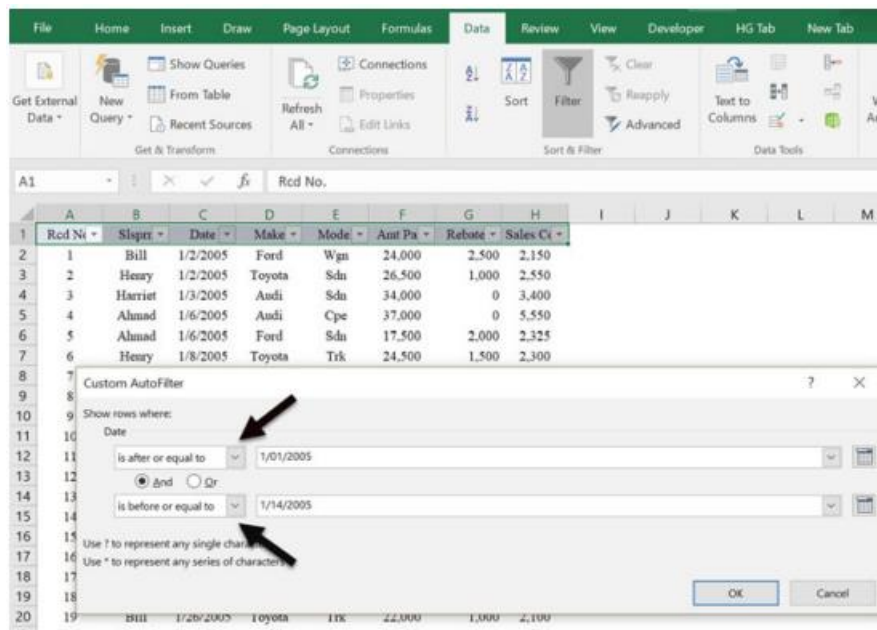
Στην συνέχεια δίνονται παραδείγματα για την χρήση του απλού και του προηγμένου φιλτραρίσματος για την κατανόησή τους.



Εικόνα 25: Filters Excel

Στην παραπάνω εικόνα εφαρμόζεται το απλό φιλτράρισμα με την μαζική επιλογή των επικεφαλίδων του δείγματός μας.

Στο επόμενο παράδειγμα είναι η χρήση του προηγμένου φίλτρου από την εντολή custom filters. Αυτή η επιλογή δίνει την δυνατότητα να δημιουργηθεί ένα ερώτημα με εύρος ημερομηνιών και όχι μόνο μία συγκεκριμένη τιμή. Γενικότερα μπορούν να εκπληρωθούν πολλές λογικές συνθήκες με τις οποίες μπορεί να κατασκευαστεί ένα ερώτημα. Υπάρχουν πολλές δυνατές επιλογές στην αναζήτηση φίλτρων όπως είναι το Equal, Before, Between κ.α..



Εικόνα 26: Advanced Filters Excel

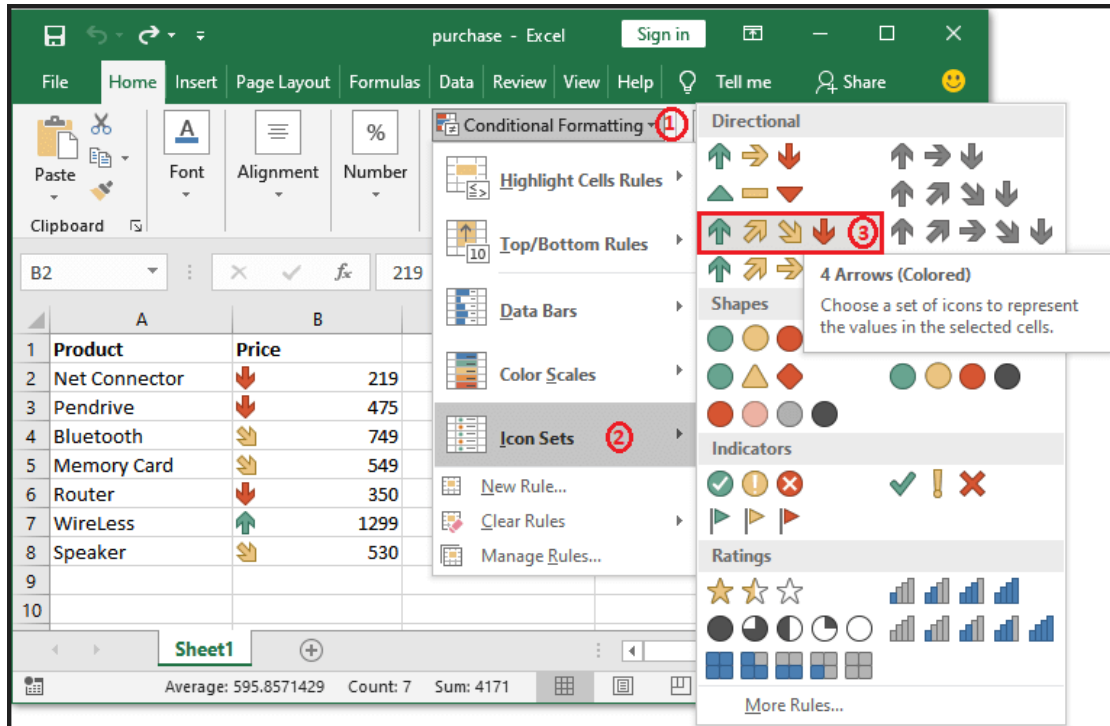
2.7.5. Οπτικοποίηση Δεδομένων

Η οπτικοποίηση δεδομένων είναι η γραφική αναπαράσταση δεδομένων και πληροφοριών. Είναι ένας τρόπος οπτικής επικοινωνίας πολύπλοκων συνόλων δεδομένων ή ιδεών, διευκολύνοντας την κατανόηση και την ερμηνεία δεδομένων. Η οπτικοποίηση δεδομένων χρησιμοποιεί γραφήματα, γραφήματα, χάρτες και άλλα οπτικά στοιχεία για την αναπαράσταση δεδομένων, τα οποία μπορούν να βοηθήσουν στον εντοπισμό μοτίβων, τάσεων και σχέσεων που μπορεί να μην είναι άμεσα εμφανείς σε έναν πίνακα ακατέργαστων δεδομένων. Η αποτελεσματική οπτικοποίηση δεδομένων μπορεί να κάνει τα δεδομένα πιο προσιτά και κατανοητά, οδηγώντας σε καλύτερη λήψη αποφάσεων και γνώσεις. (Agarwal, 2021).

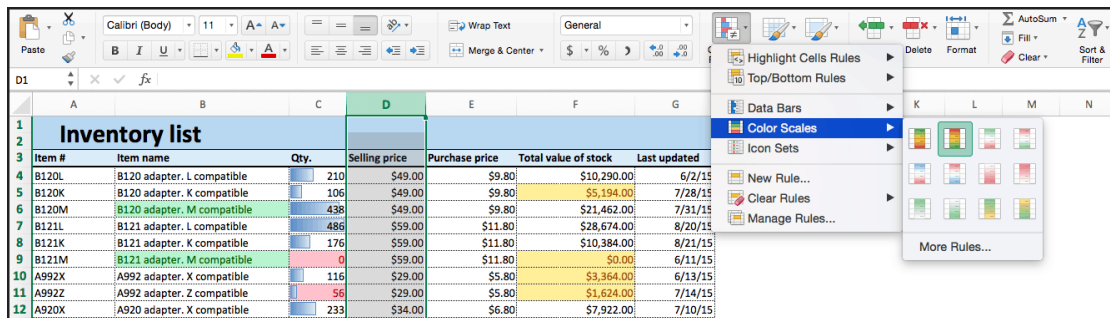
Το Excel είναι ένα ισχυρό εργαλείο για τη δημιουργία οπτικοποιημένων δεδομένων, συμπεριλαμβανομένων γραφημάτων, διαγραμμάτων και συγκεντρωτικών πινάκων. Διαθέτει ένα ευρύ φάσμα τύπων γραφημάτων για να διαλέξετε, όπως γραφήματα ράβδων, γραμμικά γραφήματα, γραφήματα πίτας, διαγράμματα διασποράς και πολλά άλλα. Για να δημιουργήσετε ένα γράφημα στο Excel, επιλέξτε τα δεδομένα που θέλετε να συμπεριλάβετε και κάντε κλικ στον τύπο γραφήματος που θέλετε να χρησιμοποιήσετε από την καρτέλα Εισαγωγή. Στη συνέχεια, μπορείτε να προσαρμόσετε το γράφημά σας χρησιμοποιώντας την καρτέλα Εργαλεία γραφήματος, η οποία σας επιτρέπει να προσαρμόσετε τη σχεδίαση, τη διάταξη και τη μορφοποίηση του γραφήματος. Το Excel σας επιτρέπει επίσης να δημιουργείτε συγκεντρωτικούς πίνακες, οι οποίοι είναι ένας ισχυρός τρόπος για τη σύνοψη και την ανάλυση μεγάλων ποσοτήτων δεδομένων.

Στην συνέχεια δίνονται παραδείγματα με κάποιους από τους πολλαπλούς τρόπους που μπορεί να επιτευχθεί η μορφοποίηση υπό όρους:

Τεχνικές Ανάλυσης Μεγάλων Δεδομένων με την χρήση Υπολογιστικών Φύλλων (Excel)



Εικόνα 27: Conditional Formatting - Icon Sets Excel



Εικόνα 28: Conditional Formatting - Color Scales Excel

Order ID	Product	Product ID	Price	Quantity	Total price	Time	Date/time
1	Veggie burger	101	\$10.00	6	\$60.00	10:40:00 AM	11/11
2	Steak sandwich	102	\$12.00	5	\$60.00	12:20:00 PM	11/22
3	Steak sandwich	102	\$12.00	3	\$36.00	2:00:00 PM	10/24
4	Steak sandwich	102	\$12.00	1	\$12.00	3:40:00 PM	10/15
5	Steak sandwich	102	\$12.00	4	\$48.00	5:20:00 PM	10/6
6	Sausage sandwich	103	\$9.00	12	\$108.00	7:00:00 PM	9/27
7	Sausage sandwich	103	\$9.00	6	\$54.00	8:40:00 PM	9/18
8	Jucy Lucy	104	\$8.00	6	\$48.00	10:20:00 PM	9/9
9	Italian beef	105	\$10.00	5	\$50.00	12:00:00 AM	8/31
10	Ham and cheese san	106	\$12.00	3	\$36.00	7:00:00 PM	8/22
11	Denver sandwich	107	\$11.00	8	\$88.00	8:40:00 PM	8/13
12	Denver sandwich	107	\$11.00	4	\$44.00	10:20:00 PM	8/4
13	Denver sandwich	107	\$11.00	2	\$22.00	12:00:00 AM	7/26
14	Crab cake sandwich	109	\$13.00	3	\$39.00	7:00:00 PM	7/17
15	Club sandwich	108	\$13.00	2	\$26.00	8:40:00 PM	7/8

Εικόνα 29: Conditional Formatting - Custom Rules Excel

2.7.6. Εργασία με πολλαπλά φύλλα εργασίας

Πολύ συχνά απαιτείται η εκτέλεση αρκετών ίδιων υπολογισμών σε άλλα φύλλα εργασίας. Το συγκεκριμένο υπολογιστικό φύλλο προσφέρει την δυνατότητα ολοκλήρωσης αυτών των υπολογισμών σε ένα φύλλο εργασίας και την εμφάνιση τους σε όλα τα άλλα χωρίς να επαναλαμβάνονται σε κάθε σελίδα (Carey & Berk, 2007).

2.7.7. What if Analysis

Η ανάλυση What-If είναι μια τεχνική μοντελοποίησης που χρησιμοποιείται στο Excel για την εξερεύνηση διαφορετικών σεναρίων αλλάζοντας τις τιμές εισόδου σε έναν τύπο φύλλου εργασίας. Βοηθά τους χρήστες να κατανοήσουν το αντίκτυπο των αλλαγών σε μία ή περισσότερες τιμές εισόδου στο αποτέλεσμα ενός τύπου. Υπάρχουν τρεις τύποι εργαλείων ανάλυσης What-If διαθέσιμα στο Excel: Goal Seek, Data Tables και Scenario Manager. (Agarwal, 2021).

- Η ανάλυση What-If είναι ένα ισχυρό εργαλείο που επιτρέπει στους χρήστες να δοκιμάσουν διαφορετικές υποθέσεις και να προβλέψουν διαφορετικά αποτελέσματα με βάση αυτές τις παραδοχές.
- Η ανάλυση What-If χρησιμοποιείται ευρέως στην επιχειρηματική και χρηματοοικονομική μοντελοποίηση, όπως η πρόβλεψη εσόδων από πωλήσεις, το κόστος του έργου και η απόδοση επένδυσης.
- Το Goal Seek είναι ένα εργαλείο που επιτρέπει στους χρήστες να ορίσουν μια τιμή στόχο για έναν τύπο και να βρουν την τιμή εισόδου που απαιτείται για την επίτευξη αυτού του στόχου. Για παράδειγμα, αν θέλετε να μάθετε πόσα πρέπει να πουλήσετε για να φτάσετε σε ένα συγκεκριμένο περιθώριο κέρδους, μπορείτε να χρησιμοποιήσετε το Goal Seek για να βρείτε την απάντηση.
- Οι πίνακες δεδομένων επιτρέπουν στους χρήστες να δημιουργήσουν έναν πίνακα που εμφανίζει τα αποτελέσματα ενός τύπου για διαφορετικές τιμές

εισόδου. Οι πίνακες δεδομένων μπορεί να είναι μονοδιάστατοι ή διδιάστατοι, ανάλογα με τον αριθμό των μεταβλητών εισόδου.

- Το Scenario Manager επιτρέπει στους χρήστες να δημιουργούν και να διαχειρίζονται διαφορετικά σενάρια με διαφορετικές τιμές εισόδου. Οι χρήστες μπορούν να συγκρίνουν τα αποτελέσματα διαφορετικών σεναρίων και να επιλέξουν το καλύτερο με βάση τους στόχους τους.
- Η ανάλυση What-If είναι ένα χρήσιμο εργαλείο για την ανάλυση κινδύνου και την ανάλυση ευαισθησίας. Με τη δοκιμή διαφορετικών σεναρίων, οι χρήστες μπορούν να εντοπίσουν τους βασικούς παράγοντες που επηρεάζουν το αποτέλεσμα και να αξιολογήσουν τον αντίκτυπο των αβεβαιοτήτων στο αποτέλεσμα.

2.8. Συμπεράσματα Κεφαλαίου

Η αξία των δεδομένων αυξάνεται καθημερινά, δίνοντας ανταγωνιστικό πλεονέκτημα στις επιχειρήσεις που βασίζονται σε αυτά. Η ανάλυση δεδομένων έχει χρησιμοποιηθεί από τις επιχειρήσεις για να αποκτήσουν δεδομένα αποτελεσματικά και να επιτύχουν εξαιρετικά αποτελέσματα (Valli, 9 C.E.). Οι αναλυτές δεδομένων υποστηρίζουν την ανάλυση δεδομένων διαχειριζόμενοι την αναλυτική προσπάθεια για την παραγωγή των καλύτερων αποτελεσμάτων με τον πιο αποτελεσματικό τρόπο. Το Microsoft Excel είναι ένα αποτελεσματικό εργαλείο για το χειρισμό των δεδομένων σε αυτήν την περίπτωση, επειδή μπορεί να χρησιμοποιηθεί για τη δημιουργία προσαρμόσιμης συγκέντρωσης δεδομένων, την οπτική απεικόνιση δεδομένων, τον υπολογισμό περιθωρίων και άλλες κοινές αναλογίες κ.λπ. Αυτό το άρθρο δίνει έμφαση στην αξία του Excel στην ανάλυση δεδομένων, τον τρόπο με τον οποίο γίνεται εφαρμόζεται και πώς ωφελεί τις επιχειρήσεις (Valli, 9 C.E.).

Το πιο προσαρμόσιμο εργαλείο που χρησιμοποιείται τώρα στις επιχειρήσεις είναι το Excel. Το πιο κρίσιμο εργαλείο δομής και ανάλυσης δεδομένων, υπάρχει από τη δεκαετία του 1980. Οι εργαζόμενοι σχεδόν σε κάθε τμήμα που μπορεί να φανταστεί κανείς, συμπεριλαμβανομένων των IT, των οικονομικών, του ανθρώπινου δυναμικού, του μάρκετινγκ, βασίζονται σε αυτό ως ζωτικής σημασίας πόρος. Ας μιλήσουμε για το πόσο πολύτιμο είναι για το πολύτιμο μάρκετινγκ μας (Data Analysis, 2022). Η διαχείριση και αποθήκευση δεδομένων έγινε με χρήση του Excel. Εξελίχτηκε σε εργαλείο για την εκτέλεση απλών υπολογισμών δεδομένων με την πάροδο του χρόνου. Μετά από μια σειρά ενημερώσεων, θεωρείται πλέον ως είσοδος στον τομέα των αναλυτικών στοιχείων. Με αυτήν την ανάρτηση, ας αγκαλιάσουμε τη δύναμη αυτού του εργαλείου και ας εισέλθουμε στον κόσμο των αναλυτικών στοιχείων μάρκετινγκ που βασίζονται στο Excel. Σε σύγκριση με τα προηγούμενα δέκα χρόνια, το Excel χρησιμοποιείται πιο συχνά σήμερα. Αυτό οφείλεται κυρίως στο γεγονός ότι οι Αναλυτές Δεδομένων το χρησιμοποιούν με τρόπο που βοηθά τον οργανισμό που υποστηρίζει την επιχείρηση. Η σημασία του Excel αυξάνεται σταδιακά αν και υπάρχουν αναδυόμενα εργαλεία όπως η Python και η R (Agarwal, 2021).

Κεφάλαιο 3: Μεθοδολογία

Το παρόν κεφάλαιο αποτελεί τον συνδετικό κρίκο του κεφαλαίου της Βιβλιογραφικής Ανασκόπησης με το κεφάλαιο της εφαρμογής που ακολουθεί και κατά συνέπεια συνδέεται έμμεσα και με τα αποτελέσματα που προκύπτουν.

Αρχικά, για να μπορέσουμε να καταλήξουμε στα αποτελέσματα, θα μελετηθούν δύο διαφορετικά παραδείγματα. Και στα δύο παραδείγματα θα χρησιμοποιηθεί πραγματικό δείγμα δεδομένων, ωστόσο πρέπει να τονίσουμε πως δεν έχει γίνει η συλλογή των παρακάτω δεδομένων, αλλά χρησιμοποιούνται έτοιμα δεδομένα τα οποία εμπιστευόμαστε την αξιοπιστία τους αλλά και την πηγή την οποία έκανε την συλλογή αυτών.

Τα δύο αυτά δείγματα παρουσιάζουν σκόπιμα διαφορετικά χαρακτηριστικά. Σκοπός αυτού είναι η εφαρμογή διαφορετικών εντολών σε κάθε περίπτωση (όπως αυτές περιγράφονται στο παραπάνω κεφάλαιο) αλλά και κοινών εντολών προκειμένου να μπορούμε να έχουμε συγκρισιμότητα μεταξύ των δύο παραδειγμάτων. Μια μεγάλη διαφορά που παρουσιάζουν τα δύο δείγματα, είναι το μέγεθός τους, το οποίο θα σταθεί καθοριστικός παράγοντας για να εξάγουμε τα αποτελέσματά μας.

Η πρώτη από τις δύο εφαρμογές, αναφέρεται σε οικονομικό παράδειγμα προσομοίωσης της συμπεριφοράς μιας κατάστασης λογαριασμού αποτελεσμάτων. Επιλέχθηκαν 500 πειραματικές παρατηρήσεις οι οποίες παρουσιάζουν είτε κέρδος είτε ζημία. Θα πραγματοποιηθεί εφαρμογή των συναρτήσεων VLOOKUP και HLOOKUP καθώς και αριθμητικών υπολογισμών με τα υπάρχοντα δεδομένα. Παράλληλα θα γίνει και η οπτικοποίηση των δεδομένων μας για να μελετηθεί η ανταπόκριση του εργαλείου πάνω σε έναν τέτοιο όγκο δεδομένων.

Η δεύτερη εφαρμογή έχει αισθητά μεγαλύτερο όγκο δεδομένων με 28.592 εγγραφές οι οποίες έχουν ως θέμα και πάλι οικονομικά στοιχεία και συγκεκριμένα αφορούν τις πωλήσεις προϊόντων μιας εταιρείας. Αντίστοιχα με το προηγούμενο παράδειγμα θα πραγματοποιηθεί εφαρμογή των συναρτήσεων VLOOKUP και HLOOKUP όπως και οι αριθμητικοί υπολογισμοί. Σκοπός αυτών είναι να αποτελέσουν τα στοιχεία σύγκρισης της λειτουργίας των συγκεκριμένων εργαλείων του excel και πως αυτό ανταποκρίνεται σε έναν πολύ μεγαλύτερο όγκο δεδομένων. Παράλληλα θα εφαρμοστούν και επιπλέον εντολές που περιγράφονται στο παραπάνω κεφάλαιο όπως η CONCATENATE αλλά και οι PivotTables όπου και είναι ιδιαίτερα χρήσιμες εντολές για την ανάλυση των Μεγάλων Δεδομένων σε πολλούς τομείς.

Τα παραπάνω χαρακτηριστικά αναφέρονται σε ποσοτικές μετρήσεις και πως μπορεί το excel να ανταποκριθεί σε αυτές. Ωστόσο εξίσου σημαντικό είναι να μελετηθούν και τα ποιοτικά χαρακτηριστικά του, τα οποία είναι κοινά ανεξαρτήτως το δείγμα μέτρησης που θα επιλέξουμε. Τα σημαντικότερα από αυτά είναι:

- ο χρόνος απόκρισης του σε κάθε μία από τις δύο περιπτώσεις
- το φιλικό προς τον χρήστη περιβάλλον και η διευκόλυνση της πλοήγησης

Ο χρόνος απόκρισης του Excel, όπως και κάθε άλλου λογισμικού, εξαρτάται από πολλούς παράγοντες όπως η ταχύτητα του υπολογιστή σας, η πολυπλοκότητα της

εργασίας που εκτελείτε, το μέγεθος του συνόλου δεδομένων και η έκδοση του Excel που χρησιμοποιείτε. Γενικά, οι νεότερες εκδόσεις του Excel τείνουν να είναι πιο γρήγορες από τις παλαιότερες εκδόσεις, καθώς είναι βελτιστοποιημένες για σύγχρονο υλικό και λειτουργικά συστήματα. Επιπλέον, η χρήση υπολογιστή με ταχύτερο επεξεργαστή και περισσότερη μνήμη μπορεί επίσης να βελτιώσει τον χρόνο απόκρισης του Excel. Η συγκεκριμένη εργασία που εκτελείτε μπορεί επίσης να επηρεάσει τον χρόνο απόκρισης του Excel. Για παράδειγμα, εάν εκτελείτε πολύπλοκους υπολογισμούς ή εργάζεστε με μεγάλα σύνολα δεδομένων, μπορεί να χρειαστεί περισσότερος χρόνος για να ανταποκριθεί το Excel από ό,τι αν εισαγάγατε απλώς κείμενο σε ένα κελί. Συνοπτικά, ο χρόνος απόκρισης του Excel επηρεάζεται από διάφορους παράγοντες και μπορεί να ποικίλλει σημαντικά ανάλογα με τις συγκεκριμένες περιστάσεις.

Η δημιουργία ενός φιλικού προς τον χρήστη περιβάλλοντος για λογισμικό είναι εξαιρετικά σημαντική. Μπορεί να έχει σημαντικό αντίκτυπο στην υιοθέτηση, τη χρησιμότητα και την αποτελεσματικότητα του λογισμικού. Ακολουθούν μερικοί λόγοι για τους οποίους ένα φιλικό προς τον χρήστη περιβάλλον είναι τόσο σημαντικό:

- Έγκριση: Ένα φιλικό προς τον χρήστη περιβάλλον διευκολύνει τους χρήστες να μάθουν και να χρησιμοποιούν το λογισμικό. Αυτό, με τη σειρά του, μπορεί να αυξήσει τα ποσοστά υιοθεσίας, καθώς οι χρήστες είναι πιο πιθανό να αγκαλιάσουν και να χρησιμοποιήσουν λογισμικό που είναι εύκολο στην πλοήγηση και κατανόηση.
- Αποδοτικότητα: Ένα φιλικό προς τον χρήστη περιβάλλον μπορεί να βελτιώσει την αποτελεσματικότητα της χρήσης λογισμικού. Όταν οι χρήστες μπορούν γρήγορα και εύκολα να έχουν πρόσβαση στις δυνατότητες και τις λειτουργίες που χρειάζονται, μπορούν να ολοκληρώσουν εργασίες πιο γρήγορα και με μεγαλύτερη ακρίβεια.
- Εκπαίδευση: Ένα φιλικό προς τον χρήστη περιβάλλον μπορεί να μειώσει τον χρόνο και τους πόρους που απαιτούνται για την εκπαίδευση. Όταν το λογισμικό είναι εύκολο στη χρήση και κατανοητό, οι χρήστες μπορούν να μάθουν να το χρησιμοποιούν πιο γρήγορα, μειώνοντας την ανάγκη για εκτεταμένες εκπαιδευτικές συνεδρίες.

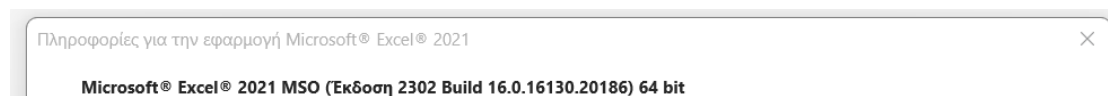
Πριν προχωρήσουμε στις εφαρμογές πρέπει να διευκρινιστούν και τα τεχνικά χαρακτηριστικά για την μελέτη των αποτελεσμάτων τους. Όπως ήδη αναφέρουμε, εξαιρετικά σημαντικό ρόλο έχει ο χρόνος απόκρισης του εργαλείου στην παρακάτω ανάλυση. Ο τρόπος με τον οποίο θα μελετηθεί είναι τα εργαλεία συγκριτικής αξιολόγησης ή ένα λογισμικό δοκιμών απόδοσης που έχει σχεδιαστεί για το Excel. Μερικά παραδείγματα τέτοιων εργαλείων περιλαμβάνουν το ExcelBench, το ExcelSpeedTest και το XLCubed.

Είναι σημαντικό να σημειωθεί ότι τα εργαλεία συγκριτικής αξιολόγησης και το λογισμικό δοκιμών απόδοσης ενδέχεται να μην είναι δωρεάν και ενδέχεται να απαιτούν τεχνική εμπειρογνομosύνη για τη ρύθμιση και χρήση. Επιπλέον, τα αποτελέσματα που λαμβάνονται ενδέχεται να μην αντικατοπτρίζουν απαραίτητα την απόδοση του Excel σε σενάρια πραγματικού κόσμου, καθώς συχνά βασίζονται σε συνθετικά ή απλουστευμένα σύνολα δεδομένων.

Σύμφωνα με τα παραπάνω καταλήξαμε στην καταγραφή του χρόνου απόκρισης σε επαναλαμβανόμενες χρήσεις για να προκύψει ένας μέσος όρος αλλά και σε συνδυασμό με άλλες μετρήσεις απόδοσης, όπως η χρήση της CPU και η χρήση μνήμης.

Για να δοθεί ολοκληρωμένη απάντηση για τα παραπάνω, πρέπει να αναφερθούμε και στα τεχνικά χαρακτηριστικά του υπολογιστή με τον οποίο έγιναν οι ακόλουθες μετρήσεις αλλά και την έκδοση του Excel την οποία χρησιμοποιήσαμε.

Σχετικά με τον υπολογιστή, χρησιμοποιήθηκε φορητός υπολογιστής (laptop) με μνήμη RAM 8 GB και CPU AMD Ryzen 5 5600 H. Η έκδοση του λειτουργικού συστήματος είναι Windows 11 Home. Για την έκδοση του Excel, στο παρακάτω απόκομμα δίνονται οι απαραίτητες πληροφορίες.



Σύμφωνα με τα παραπάνω θα ακολουθήσει το κεφάλαιο της εφαρμογής τους από όπου και θα καταλήξουμε στα τελικά μας συμπεράσματα.

Κεφάλαιο 4: Εφαρμογή του Excel για ανάλυση δεδομένων

Το συγκεκριμένο κεφάλαιο αποτελεί την εφαρμογή των παραπάνω εντολών που έχουν αναλυθεί πάνω σε πραγματικά δεδομένα προκειμένου να διαπιστώσουμε τον τρόπο ανταπόκρισης του excel σε πραγματικές εφαρμογές και την καταλληλότητα του ως εργαλείο ανάλυσης δεδομένων.

4.1. Εφαρμογή 1

Στην συγκεκριμένη περίπτωση θα εξετάσουμε ένα οικονομικό παράδειγμα που προσομοιώνει την συμπεριφορά μιας κατάστασης λογαριασμού αποτελεσμάτων. Ο πίνακας που ακολουθεί δείχνει τα στοιχεία μιας τυπικής κατάστασης εσόδων (κερδών ή ζημιών) και ο σχολιασμός των υποθέσεων που θα χρησιμοποιηθούν για τη μοντελοποίηση κάθε στοιχείου. Έχω επιλέξει μια ποικιλία κατανομών για χρήση σε αυτό το παράδειγμα, συμπεριλαμβανομένης της κανονικής κατανομής, που συχνά αναφέρεται ως καμπύλη καμπάνας λόγω του σχήματός της. Είναι μια πολύ κοινώς χρησιμοποιούμενη κατανομή που έχει ένα αξιοσημείο φάσμα εφαρμογών. Διακρίνεται από τη συμμετρία της, μια κεντρική τάση («αιχμή») και πιθανότητες για χαμηλότερες και υψηλότερες τιμές που εκτείνονται στο άπειρο, αλλά με πολύ χαμηλές πιθανότητες εμφάνισης.

Τέτοιες παρατηρήσεις μεταβλητές όπως το βάρος του ατόμου, το μέγεθος του παπουτσιού και πολλές άλλες, είναι συχνά μοντελοποιημένες με κανονικές κατανομές, αν και οι απείρως εκτεινόμενες χαμηλές και υψηλές τιμές θα μπορούσαν να οδηγήσουν σε ένα υπερβολικά μικρό μέγεθος ποδιού ή σε μέγεθος που γεμίζει ένα γήπεδο ποδοσφαίρου. Ευτυχώς, αυτές οι ακραίες τιμές έχουν πολύ, πολύ μικρή πιθανότητα εμφάνισης.

Μια κανονική κατανομή περιγράφεται από δύο παραμέτρους - τη μέση και την τυπική απόκλιση (ή διακύμανση). Ο τύπος για τη δημιουργία μιας τιμής από μια κανονική διανομή στο Excel είναι `NORMINV(RAND(), μέση, τυπική απόκλιση)`. Η συνάρτηση χρησιμοποιεί τη γνωστή συνάρτηση `RAND()` ως πρώτο όρισμα. Το `RAND()` είναι το συνάρτηση στοιχείου που λαμβάνει δείγματα από την κατανομή. Στο οικονομικό μας παράδειγμα, το ποσοστό των εσόδων που χρησιμοποιείται για τον υπολογισμό του κόστους των πωληθέντων αγαθών, `COGS`.

Τεχνικές Ανάλυσης Μεγάλων Δεδομένων με την χρήση Υπολογιστικών Φύλλων (Excel)

Component	Assumptions
Sales revenue	Sales revenue = Units * Unit Price Distributions: Discrete Units: 30%–75,000 units; 70%–10,000 units discrete unit price: 25%–\$1.50; 50%–\$2.00; 25%–\$2.50
Cost of goods sold expense (COGS)	COGS = Percentage * Sales revenue Distribution: Percentage = Normal Dist.—Mean–30; Std Dev–5
Gross margin	= Sales revenue–COGS
Variable operating expense (VOE)	VOE = Sales revenue * Percentage Distribution: Percentage = Continuous uniform Dist.—10%–20%
Contribution margin	= Gross margin – VOE
Fixed expenses (FE)	A constant value = \$6000
Operating earnings (EBIT)	= Contribution margin – Fixed expenses
Interest expense (IE)	A constant value = \$3000
Earnings before income tax	= EBIT–interest expense
Income tax expense	Conditional percentage of [EBIT–interest expense] 35% < \$20,000; 55% > = \$20,000
Net income (Profit)	= Earnings before income tax – Income tax

Εικόνα 30: Example 1 Excel (1)

Αρχικά δίνεται ο «εγκέφαλος» με τις τιμές της δειγματοληψίας αλλά και τις σταθερές τιμές που χρησιμοποιούνται στους υπολογισμούς. Στο κάτω μέρος μπορείτε να δείτε την δομή της κατάσταση κέρδους ή ζημίας που μεταφράζεται οριζόντια με κάθε σειρά να αντιπροσωπεύει και μια παρατήρηση. Εικόνα 31:

The screenshot shows an Excel spreadsheet with the following structure:

- Brain Section (Rows 1-21):** Contains input parameters for the model.
 - Sales Revenue:** cum prob, units, prob, mean, stdev.
 - COGS:** Normal distribution parameters (mean, stdev).
 - VOE:** lower limit, upper limit, prob.
 - FE:** Constant value (\$6,000).
 - IE:** Constant value (\$3,000).
 - Tax:** Rate (0.35, 0.55) and Limit (\$20,000).
- Calculations Section (Rows 22-27):** Contains formulas for each component.
 - Sales Revenue:** =VLOOKUP(RAND(), \$A4:\$B6, 2)
 - COGS:** =VLOOKUP(RAND(), \$A9:\$B12, 2)
 - VOE:** =NORMINV(RAND(), \$E\$4, \$E\$5)*\$D34/\$D35
 - Net Income (Profit):** =Earnings before Income Tax - Income Tax
- Data Table (Rows 28-42):** A table with columns: Obs., Units, Price, Revenue, COGS, VOE, FE, IE, Tax, Net Income. It shows 13 observations of the model's output.

Εικόνα 31: Example 1 Excel (2)

Από τις 13 παρατηρήσεις που είναι ορατές για το μοντέλο, μόνο 2 (παρατήρηση 8 και 10) φαίνεται να έχουν ως αποτέλεσμα ζημία (1478,45 και 303.09 δολάρια), ενώ οι υπόλοιπες παρατηρήσεις παρουσιάζουν κέρδος. Τα πλεονεκτήματα αυτής της δομής είναι:

- υπάρχουν πολλές περισσότερες σειρές ορατές σε ένα φύλλο εργασίας από τις στήλες, και
- δεδομένου ότι υπάρχουν πολλές περισσότερες σειρές σε ένα φύλλο εργασίας από στήλες, πολλές περισσότερες παρατηρήσεις είναι δυνατές.

Στο μοντέλο αυτό χρησιμοποιώ 500 πειραματικές παρατηρήσεις δήλωσης κέρδους ή ζημίας.

Οι υπολογισμοί για τον αριθμό και την τιμή που καθορίζουν τα έσοδα προέρχονται από VLOOKUP ενώ ο υπολογισμός του COGS χρησιμοποιεί τη συνάρτηση NORMINV που συζητήθηκε παραπάνω, για να καθορίσει ένα τυχαίο ποσοστό δείγματος. Το Variable Operating Expense, VOE, χρησιμοποιεί τις τιμές για κατώτερα και ανώτερα όρια (10% και 20%) μέσω γραμμικού μετασχηματισμού για να επιστρέψει συνεχείς ομοιόμορφα κατανομημένες τιμές:

$$\text{Lower limit} + (\text{upper limit} - \text{lower limit}) * \text{RAND}()$$

Ο τύπος οδηγεί σε συνεχείς τιμές μεταξύ του κατώτερου και του ανώτερου ορίου, αφού η RAND() παίρνει τιμές από 0 έως 1. Για παράδειγμα, εάν η RAND() είναι ίση με την ακραία ανώτερη τιμή 1, τότε η τιμή της έκφρασης είναι απλώς το ανώτερο όριο. Αντιστρόφως, αν η RAND() είναι ίση με την ακραία χαμηλότερη τιμή 0, τότε η έκφραση είναι ίση με το κατώτερο όριο.

Τεχνικές Ανάλυσης Μεγάλων Δεδομένων με την χρήση Υπολογιστικών Φύλλων (Excel)

Obs.	Units	Price	Revenue	COGS	VOE	FE	IE	Tax	Net Income
1	75000	\$ 2.50	\$ 187,500.00	\$ 68,476.95	\$ 19,145.83	\$ 6,000	\$ 3,000	\$ 49,982.47	\$ 40,894.75
2	10000	\$ 2.50	\$ 25,000.00	\$ 8,893.26	\$ 3,567.47	\$ 6,000	\$ 3,000	\$ 1,238.74	\$ 2,900.52
3	75000	\$ 2.00	\$ 150,000.00	\$ 39,422.32	\$ 15,408.66	\$ 6,000	\$ 3,000	\$ 47,392.96	\$ 28,776.06
4	75000	\$ 2.50	\$ 187,500.00	\$ 45,520.57	\$ 24,615.72	\$ 6,000	\$ 3,000	\$ 59,380.04	\$ 48,583.67
5	75000	\$ 2.00	\$ 150,000.00	\$ 43,366.32	\$ 21,534.48	\$ 6,000	\$ 3,000	\$ 42,654.56	\$ 34,244.64
6	10000	\$ 2.50	\$ 25,000.00	\$ 7,652.84	\$ 3,029.40	\$ 6,000	\$ 3,000	\$ 1,923.83	\$ 3,572.83
7	10000	\$ 2.00	\$ 20,000.00	\$ 7,652.84	\$ 3,029.40	\$ 6,000	\$ 3,000	\$ 1,327.08	\$ 2,464.58
8	10000	\$ 1.50	\$ 15,000.00	\$ 7,652.84	\$ 3,029.40	\$ 6,000	\$ 3,000	\$ -	\$ (1,478.45)
9	10000	\$ 2.00	\$ 20,000.00	\$ 7,652.84	\$ 3,029.40	\$ 6,000	\$ 3,000	\$ 897.30	\$ 1,665.41
494	10000	\$ 1.50	\$ 15,000.00	\$ 4,845.68	\$ 1,573.01	\$ 6,000	\$ 3,000	\$ -	\$ (418.48)
495	10000	\$ 2.00	\$ 20,000.00	\$ 6,798.79	\$ 2,437.13	\$ 6,000	\$ 3,000	\$ 617.43	\$ 1,146.65
496	10000	\$ 1.50	\$ 15,000.00	\$ 4,122.93	\$ 1,594.46	\$ 6,000	\$ 3,000	\$ 98.91	\$ 183.70
497	75000	\$ 1.50	\$ 112,500.00	\$ 30,274.33	\$ 15,744.23	\$ 6,000	\$ 3,000	\$ 31,064.79	\$ 25,416.64
498	10000	\$ 1.50	\$ 15,000.00	\$ 3,515.26	\$ 1,905.84	\$ 6,000	\$ 3,000	\$ 202.61	\$ 376.28
499	10000	\$ 2.00	\$ 20,000.00	\$ 6,575.12	\$ 2,268.64	\$ 6,000	\$ 3,000	\$ 754.68	\$ 1,401.55
500	10000	\$ 2.50	\$ 25,000.00	\$ 7,656.26	\$ 3,548.29	\$ 6,000	\$ 3,000	\$ 1,678.41	\$ 3,117.06
max									\$ 51,664.36
min									\$ (2,930.71)
avg									\$ 10,863.95
stdev									\$ 15,625.46
loss count									75
profit count									425

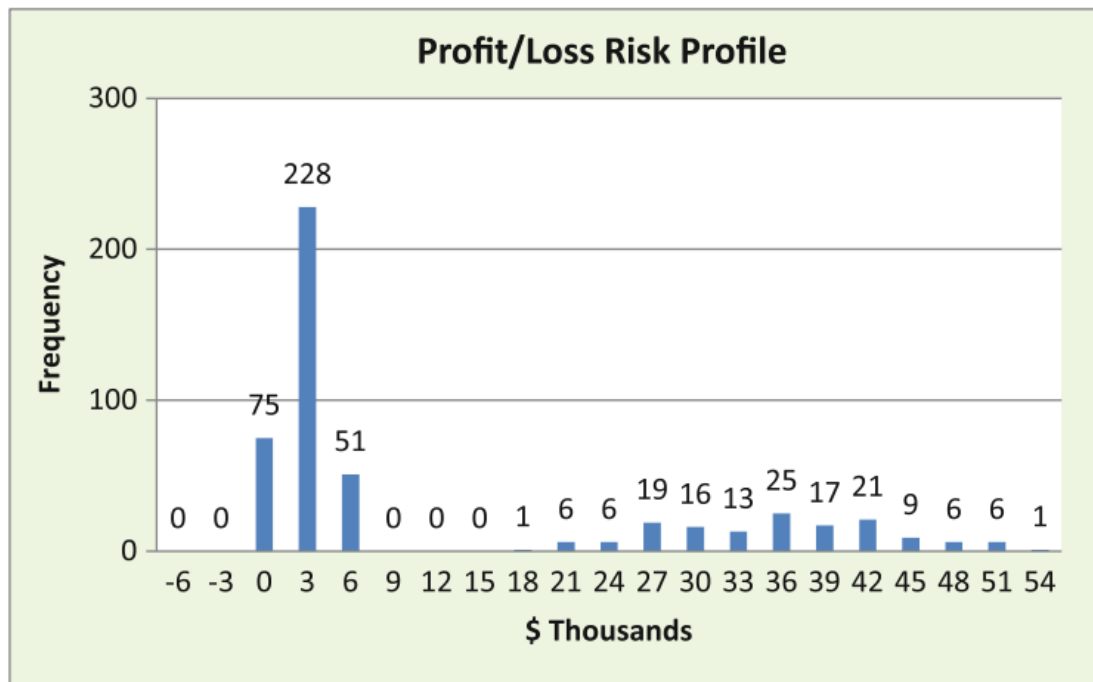
Εικόνα 32: Example 1 Excel (3)

Στην παραπάνω εικόνα βλέπουμε συνοπτικά στατιστικά στοιχεία (εύρος κελιών I530:J535) για την προσομοίωση από τις 500 παρατηρήσεις. Το μέσο κέρδος ή ζημιά είναι θετικό, οδηγώντας σε κέρδος ύψους 10.863,95 δολάρια και μια πολύ σημαντική απόκλιση 15.625,46 δολαρίων. Ο συντελεστής της διακύμανσης για τις 500 παρατηρήσεις είναι μεγαλύτερος από την μονάδα (15.625,46/10.863,95). Και έτσι, το μοντέλο μας δείχνει ένα πολύ μεγάλο εύρος πιθανών τιμών με μέγιστη τα +51.957,41 \$ και ελάχιστη τα -2930,71 \$. Επίσης πρέπει να παρατηρήσουμε το κελί J534 που έχει την τιμή 75, συμβολίζει τις αρνητικές παρατηρήσεις ή το 15% των συνολικών αποτελεσμάτων.

Ας εξετάσουμε την σειρά 30 του φύλλου εργασίας όπου είναι και η πρώτη παρατήρηση. Ο αριθμός των παραγόμενων μονάδων είναι 75.000 με την τιμή ανά μονάδα να διαμορφώνεται στα 2.50 \$. Τα έσοδα προκύπτουν με την χρήση VLOOKUP. Το κόστος πωληθέντων αγαθών (COGS), 68.476,95 \$, είναι ένα ποσοστό των εσόδων και προσδιορίζεται από μια τιμή αποτελέσματος κανονικής κατανομής διαιρούμενη με το 100. Η μεταβλητή για το λειτουργικό κόστος (VOE) προσδιορίζεται με δειγματοληψία μιας Ομοιόμορφης κατανομής (19.145,83 \$). Τα πάγια έξοδα και τόκοι (FE και IE) είναι σταθερές τιμές των \$6000 και \$3000, αντίστοιχα. Τέλος, ο φόρος καθορίζεται από τη λογική - εάν υπάρχει απώλεια ο φορολογικός συντελεστής είναι μηδέν και εάν προκύψει κέρδος, καθορίστε το επίπεδο αυτού και εφαρμόστε το κατάλληλο φορολογικό συντελεστή -. Ο υπολογισμός του καθαρού εισοδήματος είναι

το άθροισμα όλων των εξόδων αφαιρείται από τα έσοδα και καταλήγει σε κέρδος 40.894,75 \$.

Στην συνέχεια δίνεται και η διαγραμματική απεικόνιση του ρίσκου που υπάρχει και την συχνότητα εμφάνισης σύμφωνα με το δείγμα των παραπάνω 500 περιπτώσεων. Παρατηρούμε ότι όλες οι απώλειες είναι μικρής κλίμακας (0-2.999 δολάρια) και των μεγαλύτερο μέρος των κερδών είναι επίσης μικρής κλίμακας (1 – 8.999 δολάρια). Όσο αυξάνεται το κέρδος πλησιάζουμε και στις ακραίες παρατηρήσεις του δείγματός μας.



Εικόνα 33: Example 1 Excel (4)

4.2. Εφαρμογή 2

Αντίστοιχα με την προηγούμενη περίπτωση, έτσι και σε αυτήν το δείγμα θα αποτελέσουν οικονομικά δεδομένα, τα οποία φαίνονται στην εικόνα παρακάτω. Υπάρχουν οι πωλήσεις προϊόντων εταιρίας και δίνονται αναλυτικά στοιχεία για αυτές (η χώρα, η ημερομηνία συναλλαγής, το κόστος παραγωγής, οι πωλήσεις, το κέρδος, κα.). Συγκεκριμένα σε αυτό το φύλλο εργασίας έχουμε 28592 εγγραφές, αρκετές περισσότερες σε σχέση με το προηγούμενο παράδειγμα και αριθμό ικανό για να καταλήξουμε στα συμπεράσματά μας.

Τεχνικές Ανάλυσης Μεγάλων Δεδομένων με την χρήση Υπολογιστικών Φύλλων (Excel)

Region	Country	Division	Units Sold	Manufacturing Costs	Ramp-Up Costs	R&D Costs	Marketing Costs	SG&A Costs	COGS	Profit	Month	Month Name	Year
Government	Canada	Carrtera	None	1618.5	3,00	20,00	32,370.00	-	32,370.00	16,185.00	1/1/2014	1 January	2014
Government	Germany	Carrtera	None	1321	3,00	20,00	26,420.00	-	26,420.00	13,210.00	1/1/2014	1 January	2014
Midmarket	France	Carrtera	None	2178	3,00	15,00	32,670.00	-	32,670.00	21,780.00	1/6/2014	6 June	2014
Midmarket	Germany	Carrtera	None	488	3,00	15,00	13,320.00	-	13,320.00	8,880.00	1/6/2014	6 June	2014
Midmarket	Mexico	Carrtera	None	2470	3,00	15,00	37,050.00	-	37,050.00	24,700.00	1/6/2014	6 June	2014
Government	Germany	Carrtera	None	1513	3,00	350,00	529,530.00	-	529,530.00	899,580.00	1/12/2014	12 December	2014

Εικόνα 34: Example 2 Excel (1)

Αρχικά έγινε η εφαρμογή των πιο συνηθισμένων εντολών και μερικών αριθμητικών υπολογισμών. Στην πρώτη εφαρμογή όπως μπορούμε να δούμε στην παρακάτω εικόνα, έγινε η εφαρμογή της εντολής CONCATENATE για να μπροσέσουμε να βάλουμε σε μία κοινή στήλη τον μήνα και το έτος, τα οποία αφορά η συναλλαγή και η σύνταξη της εντολής αλλά και τα αποτελέσματα αυτής τα βλέπουμε στην στήλη Q του φύλλου εργασίας. Ταυτόχρονα εφαρμόστηκε και το Conditional Formatting με σκοπό να ξεχωρίζουν οι συναλλαγές στις οποίες προκύπτει ζημία. Το Conditional Formatting εφαρμόστηκε αναζητώντας στην στήλη L, όλες τις τιμές που περικλείονται από παρενθέςεις (...), όπου και είναι τρόπος με την οποίο δηλώνονται οι ζημίες στο συγκεκριμένο φύλλο εργασίας.

Region	Country	Division	Units Sold	Manufacturing Costs	Ramp-Up Costs	R&D Costs	Marketing Costs	SG&A Costs	COGS	Profit	Month	Month Name	Year
Government	Canada	Carrtera	None	1618.5	3,00	20,00	32,370.00	-	32,370.00	16,185.00	1/1/2014	1 January	2014
Government	Germany	Carrtera	None	1321	3,00	20,00	26,420.00	-	26,420.00	13,210.00	1/1/2014	1 January	2014
Midmarket	France	Carrtera	None	2178	3,00	15,00	32,670.00	-	32,670.00	21,780.00	1/6/2014	6 June	2014
Government	France	Medium	2487	3,00	7,00	17,420.00	870.25	15,530.00	122.455	3,218.25	10/2/2014	12 October	2014
Government	France	Medium	1384.5	5,00	350,00	484,575.00	24,228.75	466,346.25	359,970.00	100,376.25	1/1/2014	1 January	2014
Enterprise	United States of America	Medium	3627	5,00	125,00	433,375.00	22,668.75	430,706.25	433,240.00	(4,533.75)	7/7/2014	7 July	2014

Εικόνα 35: Example 2 Excel (2)

Στην συνέχεια έγινε η εφαρμογή απλών υπολογισμών όπως είναι η συνάρτηση Average και SUM. Η αναπόκριση του excel ήταν αρκετά γρήγορη σε σχέση με τις δύο παραπάνω εντολές που παρουσιάστηκαν. Τα αποτελέσματα από τους μαθηματικούς υπολογισμούς δίνονται στην παρακάτω εικόνα.

Τεχνικές Ανάλυσης Μεγάλων Δεδομένων με την χρήση Υπολογιστικών Φύλλων (Excel)

28578	Enterprise	France	Velo	663	\$	120,00	\$	125,00	\$	82.875,00	\$	12.431,25	\$	70.443,75	\$	79.560,00	\$	(9.116,25)	September	2014	September, 2014
28579	Small Business	United States of Ai	Velo	2574	\$	120,00	\$	300,00	\$	772.200,00	\$	\$115.830,00	\$	656.370,00	\$	643.500,00	\$	12.870,00	November	2013	November, 2013
28580	Enterprise	United States of Ai	Velo	2438	\$	120,00	\$	125,00	\$	304.750,00	\$	45.712,50	\$	259.037,50	\$	292.560,00	\$	(33.522,50)	December	2013	December, 2013
28581	Channel Partners	United States of Ai	Velo	914	\$	120,00	\$	12,00	\$	10.968,00	\$	1.645,20	\$	9.322,80	\$	2.742,00	\$	6.580,80	December	2014	December, 2014
28582	Government	Canada	VTT	865,5	\$	250,00	\$	20,00	\$	17.310,00	\$	2.596,50	\$	14.713,50	\$	8.655,00	\$	6.058,50	July	2014	July, 2014
28583	Midmarket	Germany	VTT	492	\$	250,00	\$	15,00	\$	7.380,00	\$	1.107,00	\$	6.273,00	\$	4.920,00	\$	1.353,00	July	2014	July, 2014
28584	Government	United States of Ai	VTT	267	\$	250,00	\$	20,00	\$	5.340,00	\$	801,00	\$	4.539,00	\$	2.670,00	\$	1.869,00	October	2013	October, 2013
28585	Midmarket	Germany	VTT	1175	\$	250,00	\$	15,00	\$	17.625,00	\$	2.643,75	\$	14.981,25	\$	11.750,00	\$	3.231,25	October	2014	October, 2014
28586	Enterprise	Canada	VTT	2954	\$	250,00	\$	125,00	\$	369.250,00	\$	55.387,50	\$	313.862,50	\$	354.480,00	\$	(40.617,50)	November	2013	November, 2013
28587	Enterprise	Germany	VTT	552	\$	250,00	\$	125,00	\$	69.000,00	\$	10.350,00	\$	58.650,00	\$	66.240,00	\$	(7.590,00)	November	2014	November, 2014
28588	Government	France	VTT	293	\$	250,00	\$	20,00	\$	5.860,00	\$	879,00	\$	4.981,00	\$	2.930,00	\$	2.051,00	December	2014	December, 2014
28589	Small Business	France	Amarilla	2475	\$	260,00	\$	300,00	\$	742.500,00	\$	\$111.375,00	\$	631.125,00	\$	618.750,00	\$	12.375,00	March	2014	March, 2014
28590	Small Business	Mexico	Amarilla	546	\$	260,00	\$	300,00	\$	163.800,00	\$	24.570,00	\$	139.230,00	\$	136.500,00	\$	2.730,00	October	2014	October, 2014
28591	Government	Mexico	Montana	1368	\$	5,00	\$	7,00	\$	9.576,00	\$	1.436,40	\$	8.139,60	\$	6.840,00	\$	1.299,60	February	2014	February, 2014
28592	Government	Canada	Paseo	723	\$	10,00	\$	7,00	\$	5.061,00	\$	759,15	\$	4.301,85	\$	3.615,00	\$	686,85	April	2014	April, 2014
28593	Channel Partners	United States of Ai	VTT	1806	\$	250,00	\$	12,00	\$	21.672,00	\$	3.250,80	\$	18.421,20	\$	5.418,00	\$	13.003,20	May	2014	May, 2014
28594					\$	96,25				\$	4.813.105,96671		\$	4.136.687.309,50		\$	676.418.677,21				
28595			Average																		
28596			=AVERAGE(F1:F28593)																		
28597																					
28598																					
28599																					
28600																					
28601																					
28602																					
28603																					
28604																					
28605																					
28606																					
28607																					
28608																					
28609																					
28610																					
28611																					
28612																					
28613																					
28614																					
28615																					

Εικόνα 36: Εφαρμογή 2 Excel (3)

Τέλος, εφαρμόστηκαν και οι πιο εξειδικευμένες και χρήσιμες ταυτόχρονα, εντολές στην ανάλυση δεδομένων όπως είναι η VLOOK και PivotTables. Η δομή αυτών και τα αποτελέσματα που προκύπτουν δίνονται στις εικόνες που ακολουθούν:

Country	Capital
Canada	Ottawa
France	Paris
Germany	Berlin
Mexico	Mexico City
United States of America	Washington

Row Labels	Sum of Units Sold	Sum of Profit
2013	10743003	154991943,4
2014	35170781,5	521427613,8
Grand Total	45913786,5	676418677,2

Εικόνα 37: Εφαρμογή 2 Excel (4)

Αρχικά, παρατηρούμε στο κάτω αριστερό μέρος της παραπάνω εικόνας την εφαρμογή του PivotTable με τα αποτελέσματα που δίνει για τις πωλήσεις αλλά και το κέρδος που προκύπτει για κάθε μια από τις χώρες αυτές για τα έτη 2013 και 2014. Ακριβώς δίπλα φαίνεται ο πίνακας που κατασκευάστηκε για να αποτελέσει το Lookup Value, δίνοντας μας, της πρωτεύουσα για κάθε μία από τις χώρες που εμφανίζονται στα δεδομένα μας. Η σύνταξη της εντολής παρουσιάζεται παρακάτω, μαζί με τα δεδομένα που προέκυψαν από αυτήν.

Τεχνικές Ανάλυσης Μεγάλων Δεδομένων με την χρήση Υπολογιστικών Φύλλων (Excel)

The screenshot shows an Excel spreadsheet with a data table. The columns are labeled A through V. The data table starts at row 2 and ends at row 55. The columns are: A: Company, B: Country, C: Brand, D: Units Sold, E: Manufacturing Price, F: Sales Price, G: Gross Sales, H: Discounts, I: Sales, J: COGS, K: Profits, L: Month, M: Home. The data rows contain various company names, countries, brands, and financial figures. The spreadsheet also shows the ribbon at the top with tabs for File, Home, Insert, Page Layout, Formulas, Data, Review, View, and Help. The status bar at the bottom indicates the active cell is R2 and contains the formula =VLOOKUP(B2:F226011:55226052;TRUE).

Εικόνα 38: Εφαρμογή 2 Excel (5)

Στην στήλη R (Capital), δίνονται τα αποτελέσματα για τις πρωτεύουσες των χωρών αλλά φαίνεται και η σύνταξη της εντολής. Στην συγκεκριμένη εντολή παρατηρούμε την χρήση του συμβόλου του δολαρίου (\$). Μας δίνει την δυνατότητα να κρατήσουμε σταθερό την απόλυτη θέση ενός κελιού όσον αφορά και τον αριθμό γραμμής και το γράμμα στήλης. Στην συγκεκριμένη περίπτωση θέλουμε να κρατάμε σταθερές και τις δύο επιλογές καθώς, ο πίνακας που χρησιμοποιούμε για να εξάγουμε τα αποτελέσματα από την VLOOKUP, είναι σε συγκεκριμένη θέση στο φύλλο εργασίας όπως έχουμε δείξει και παραπάνω.

Αυτές οι δύο εφαρμογές στην παρούσα ενότητα, χρησιμοποιήθηκαν για να καταλήξουμε στο τελευταίο κεφάλαιο της Διπλωματικής Εργασίας, όπου και απαντάται το ερευνητικό ερώτημα που είχε τεθεί στο Κεφάλαιο της Εισαγωγής.

4.3. Συνδυαστικά χαρακτηριστικά των δύο εφαρμογών

Πέρα από τα ποσοτικά χαρακτηριστικά που μελετιούνται παραπάνω, υπάρχουν και τα ποιοτικά χαρακτηριστικά του excel. Ένα από τα μεγάλα θετικά που παρουσιάζει είναι η ευκολία στη χρήση του. Ακόμη και ένας χρήστης που δεν είναι οικείος με το συγκεκριμένο εργαλείο, ο τρόπος με τον οποίο παρουσιάζονται οι πιο συνηθισμένες λειτουργίες του, το κάνουν πολύ εύκολα προσβάσιμο στον απλό χρήστη. Ακόμα και στην ανάπτυξη συναρτήσεων, όπου κάποιος χρήστης ακολουθεί τα βήματα που προσφέρει το excel και δεν γράφει μόνος του την σύνταξη της συνάρτησης, είναι εύκολα κατανοητά ώστε να πετύχει το επιθυμητό αποτέλεσμα.

Αντίστοιχα με τα παραπάνω η οπτικοποίηση των δεδομένων, γίνεται επίσης με πολύ εύκολο τρόπο και επιτρέπει σε έναν χρήστη να παράγει γραφήματα (απλού τύπου) ανεξαρτήτως από το επίπεδο εξοικείωσής του. Παράλληλα με την εύκολη λειτουργία του, τα αρχεία του excel είναι διαθέσιμα για επεξεργασία σε Windows, Mac, IOS και android.

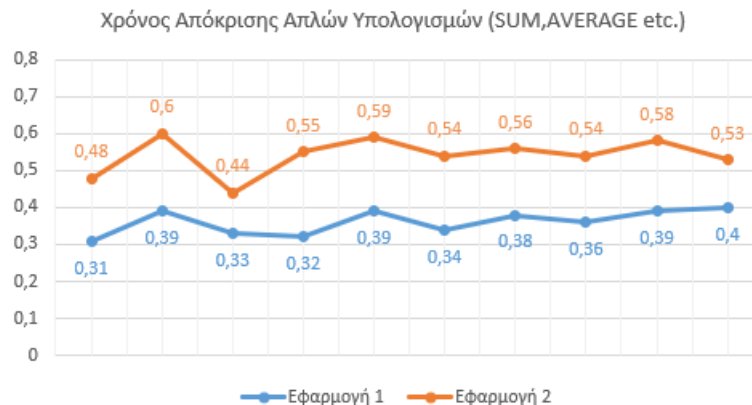
Ωστόσο σε πολύ μεγάλους όγκους δεδομένων το συγκεκριμένο εργαλείο δεν μπορεί να ανταπεξέλθει αντίστοιχα με γλώσσες προγραμματισμού που κάνουν την ανάλυση

δεδομένων πολύ πιο γρήγορη. Παράλληλα με αυτό παρατηρήθηκε πως το Excel μπορεί να αποθηκεύσει μόνο 1 εκατομμύριο σειρές δεδομένων, κάτι που μπορεί να μην είναι αρκετό για ορισμένους χρήστες και ειδικότερα όταν πρόκειται για επαγγελματική χρήση του.

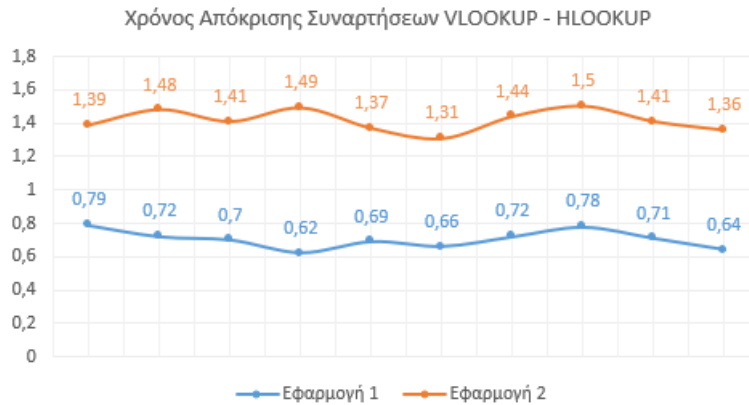
4.3.1. Τεχνικά Χαρακτηριστικά

Όπως αναφέρεται και στο προηγούμενο κεφάλαιο ο καλύτερος διαθέσιμος για να προκύψουν τα τεχνικά χαρακτηριστικά του εργαλείου, είναι η εκτέλεση των παραπάνω υπολογισμών πολλαπλές φορές με σκοπό την απόκτηση ενός μέσο όρου στον χρόνο απόκρισης. Παράλληλα θα γίνει και η μελέτης χρήσης του υπολογιστή σχετικά με την μνήμη RAM αλλά και τον επεξεργαστή.

Στην συνέχεια παρουσιάζονται τα αποτελέσματα μέτρησης για την ταχύτητα ανταπόκρισης του εργαλείου στην εκτέλεση των παραπάνω εντολών. Πρέπει να αναφέρουμε πως οι υπολογισμοί γίνανε 10 φορές με σκοπό να προκύψει ένας μέσος όρος για τον χρόνο απόκρισης και τα αποτελέσματα να ανταποκρίνεται όσο καλύτερα γίνεται στην πραγματικότητα. Στο πρώτο από τα δύο γραφήματα βλέπουμε τους χρόνους που προκύψανε στην εκτέλεση απλών υπολογιστικών πράξεων. Με το μπλε χρώμα παρουσιάζεται ο απαιτούμενος χρόνος σε δευτερόλεπτα για την εφαρμογή των υπολογισμών Εφαρμογή 1 με τις 500 εγγραφές ενώ με το πορτοκάλι χρώμα η ανταπόκριση στην Εφαρμογή 2 που αποτελεί και πιο σύνθετο παράδειγμα.



Στο δεύτερο γράφημα φαίνονται και πάλι οι χρόνοι απόκρισης του εργαλείου, αυτή τη φορά με την χρήση συναρτήσεων και όχι με απλούς υπολογισμούς. Και πάλι οι χρόνοι που αναγράφονται αναφέρονται σε δευτερόλεπτα.

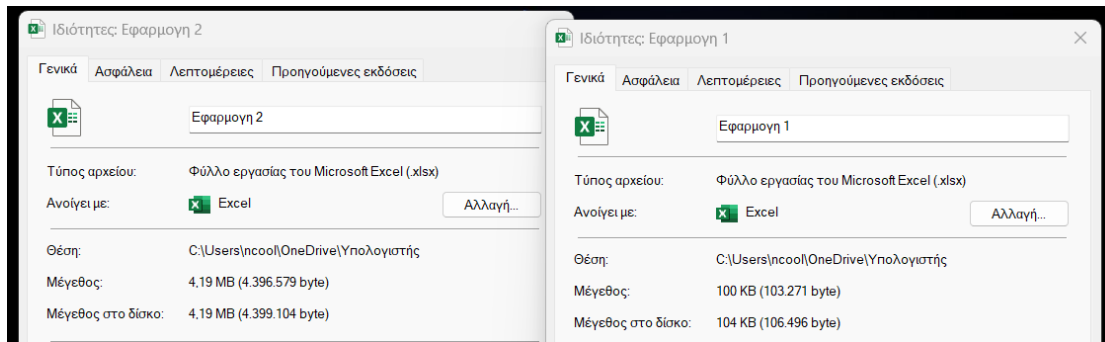


Στα συμπεράσματα των δύο παραπάνω παραδειγμάτων καταλήγουμε πως σίγουρα, όσο αυξάνεται ο όγκος των δεδομένων που απαιτούνται για μια πράξη ή συνάρτηση, ο χρόνος απόκρισης μεταβάλλεται. Ωστόσο στους πιο σύνθετους υπολογισμούς, το excel παρατηρούνται μεγαλύτερες αποκλίσεις αναλογικά με αυτές των απλών πράξεων, σύμφωνα πάντα με τον όγκο των δεδομένων.

Ωστόσο πρέπει να αναφερθεί πως όλοι οι παραπάνω υπολογισμοί εκτελέστηκαν βάση την ανθρώπινη αντίδραση και όχι κάποιου εργαλείου ακριβείας που λειτουργεί με αυτοματισμούς. Για αυτό και επιλέχθηκε να πραγματοποιηθούν 10 φορές οι υπολογισμοί, έτσι ώστε το δείγμα μας να είναι όσο το δυνατόν πιο αντιπροσωπευτικό.

Σχετικά με την μνήμη RAM, χωρίς να εκτελούνται παράλληλα σύνθετες διαδικασίες στον υπολογιστή παρά μόνο κάποιες απλές διεργασίες στην πλοήγηση στο διαδίκτυο, παρατηρήθηκαν υψηλές τιμές κάλυψης μνήμης. Αυτές οι τιμές και στα δύο παραδείγματα, κατά την εκτέλεση των υπολογισμών ήταν άνω του 85%. Σύμφωνα με τις απαιτήσεις που ολοένα και αυξάνονται, σε ότι αφορά τα τεχνικά χαρακτηριστικά των υπολογιστών, πρέπει να επισημάνουμε πως τα 8GB μνήμης RAM δεν αποτελούν και την καλύτερη δυνατή επιλογή. Ωστόσο ότι απαιτείται τόσο μεγάλη απαίτηση, για την εκτέλεση των υπολογισμών, σίγουρα δεν αποτελεί θετικό χαρακτηριστικό για το excel. Αντιθέτως, σε κανένα από τα δύο παραδείγματα δεν ήταν απαραίτητη όλη η επεξεργαστική δύναμη του υπολογιστή, και οι τιμές που παρατηρήθηκαν ήταν κάτω του 15% σε όλες τις περιπτώσεις, με τις διαφορές να είναι μη υπολογίσιμες.

Σύμφωνα με τον όγκο μνήμης που καλύπτει ένα αρχείο στον υπολογιστή, στην ακόλουθη εικόνα παρουσιάζονται το μέγεθος των αρχείων που προκύπτουν στο τέλος των δύο παραπάνω εφαρμογών:



Εικόνα 39: Τεχνικά Χαρακτηριστικά

Είναι χαρακτηριστική η διαφορά στο μέγεθος των δύο αρχείων, με αυτό της εφαρμογής 2 να είναι υπερ. πολλαπλάσιο από αυτό της πρώτης εφαρμογής. Συγκεκριμένα το ένα αρχείο βλέπουμε ότι είναι περίπου 100KB ενώ το δεύτερο 4.2MB. Ο αριθμός των εγγραφών στα δύο παραδείγματα ήταν 500 και 28600 περίπου. Παρατηρούμε ότι ο όγκος του αρχείου δεν αυξάνεται ανάλογα με τον όγκο των εγγραφών, αλλά το αρχείο γίνεται οριακά πιο «ελαφρύ» σε σύγκριση με το προηγούμενο καθώς με την απλή αναλογική, σύμφωνα με το παράδειγμα 1, ο όγκος του παραγόμενου αρχείου από το παράδειγμα 2 θα έπρεπε να είναι 5.6MB.

4.3.2. Ποιοτικά Χαρακτηριστικά

Το λογισμικό έχει μια οικεία περιβάλλον με μια σειρά από χαρακτηριστικά που είναι εύκολη στην πρόσβαση και τη χρήση. Ακολουθούν ορισμένοι λόγοι για τους οποίους το Excel θεωρείται φιλικό προς τον χρήστη:

- **Οπτικά βοηθήματα:** Το Excel παρέχει στους χρήστες οπτικά βοηθήματα, όπως χρωματικά κωδικοποιημένα κελιά και εικονίδια, για να βοηθούν τους χρήστες να αναγνωρίζουν εύκολα διαφορετικές δυνατότητες και λειτουργίες του λογισμικού.
- **Ribbon Menu:** Το Ribbon Menu του Excel διευκολύνει τους χρήστες να βρίσκουν και να χρησιμοποιούν διάφορες λειτουργίες και λειτουργίες. Το μενού Ribbon ομαδοποιεί παρόμοιες λειτουργίες, καθιστώντας ευκολότερο τον εντοπισμό και τη χρήση των συγκεκριμένων εργαλείων που απαιτούνται για μια εργασία.
- **Δυνατότητες βοήθειας:** Το Excel διαθέτει ένα εκτεταμένο σύστημα βοήθειας που παρέχει στους χρήστες πρόσβαση σε λεπτομερή εκπαιδευτικά προγράμματα και υλικό αναφοράς. Αυτό μπορεί να είναι ιδιαίτερα χρήσιμο για νέους χρήστες που εξακολουθούν να μαθαίνουν το λογισμικό.

Αυτά αποτελούν μερικά από τα σημαντικότερα ποιοτικά χαρακτηριστικά που παρατηρήθηκαν και στην συνέχεια, με αφορμή όλα τα παραπάνω οδηγούμαστε στο επόμενο κεφάλαιο όπου και έχουμε τα συμπεράσματα που προκύπτουν.

Κεφάλαιο 5: Συμπεράσματα

Το Excel είναι ένα ισχυρό εργαλείο για τη διαχείριση και την ανάλυση δεδομένων, αλλά έχει επίσης ορισμένα μειονεκτήματα που μπορεί να περιορίσουν τη χρησιμότητά του για ορισμένες εργασίες. Ακολουθούν μερικά από τα κοινά μειονεκτήματα του Excel τα οποία εντοπίζονται από την Βιβλιογραφική ανασκόπηση αλλά και από την εφαρμογή που προηγήθηκε στο παραπάνω κεφάλαιο:

1. Περιορισμένη χωρητικότητα δεδομένων - Το Excel έχει περιορισμένη χωρητικότητα για αποθήκευση και ανάλυση δεδομένων. Καθώς τα σύνολα δεδομένων μεγαλώνουν, το Excel μπορεί να γίνει αργό και επιρρεπές σε σφάλματα.
2. Έλλειψη Διακυβέρνησης Δεδομένων - Το Excel δεν παρέχει ισχυρές δυνατότητες διακυβέρνησης δεδομένων, γεγονός που μπορεί να οδηγήσει σε ασυνέπειες, σφάλματα και παραβιάσεις δεδομένων. Οι χρήστες μπορούν επίσης να τροποποιήσουν δεδομένα χωρίς σαφή διαδρομή ελέγχου.
3. Περιορισμένες δυνατότητες συνεργασίας - Αν και το Excel διαθέτει ορισμένες δυνατότητες συνεργασίας, όπως κοινόχρηστα βιβλία εργασίας, δεν είναι μια ολοκληρωμένη πλατφόρμα συνεργασίας. Μπορεί να είναι δύσκολο για πολλούς χρήστες να εργαστούν στο ίδιο βιβλίο εργασίας ταυτόχρονα και δεν υπάρχει έλεγχος έκδοσης.
4. Περιορισμένες δυνατότητες αναφοράς - Αν και το Excel διαθέτει ορισμένες βασικές δυνατότητες αναφοράς, ενδέχεται να μην επαρκεί για περίπλοκες ανάγκες αναφοράς, ειδικά όταν πρόκειται για δυναμικές και αυτοματοποιημένες αναφορές.
5. Επιρρεπής σε σφάλματα - Το Excel είναι ένα ισχυρό εργαλείο, αλλά είναι επίσης επιρρεπές σε σφάλματα. Η ανακριβής εισαγωγή δεδομένων, τα σφάλματα τύπου και άλλα λάθη μπορούν όλα να οδηγήσουν σε εσφαλμένα αποτελέσματα.
6. Κίνδυνοι ασφαλείας - Τα αρχεία Excel μπορεί να είναι ευάλωτα σε κινδύνους ασφαλείας, όπως κακόβουλο λογισμικό και μη εξουσιοδοτημένη πρόσβαση.
7. Περιορισμένη ενσωμάτωση με άλλα εργαλεία - Το Excel ενδέχεται να μην ενσωματώνεται καλά με άλλα εργαλεία στον οργανισμό σας, γεγονός που μπορεί να οδηγήσει σε διπλή προσπάθεια και αναποτελεσματικότητα.

Συνολικά, το Excel είναι ένα ισχυρό εργαλείο για τη διαχείριση και την ανάλυση δεδομένων, αλλά έχει επίσης ορισμένους περιορισμούς που πρέπει να λαμβάνονται υπόψη κατά τη χρήση του για τις επιχειρηματικές εργασίες.

Για τα Τεχνικά Χαρακτηριστικά:

Ο χρόνος απόκρισης διαπιστώθηκε ότι είναι και από τα σημαντικότερα προβλήματα που αντιμετωπίζει ο χρήστης με την σταδιακή αύξηση των δεδομένων. Συγκεκριμένα, στο Excel ο χρόνος απόκρισης για εργασίες ανάλυσης δεδομένων μπορεί να ποικίλλει ανάλογα με την πολυπλοκότητα και το μέγεθος των δεδομένων που αναλύονται, την ισχύ του υπολογιστή που εκτελεί την ανάλυση και την αποτελεσματικότητα των τεχνικών ανάλυσης που χρησιμοποιούνται. Οι απλές αναλύσεις δεδομένων μπορούν συχνά να εκτελεστούν γρήγορα στο Excel, ενώ οι πιο σύνθετες αναλύσεις που περιλαμβάνουν μεγάλα σύνολα δεδομένων ή πολύπλοκους αλγόριθμους ενδέχεται να χρειαστούν περισσότερο χρόνο για να ολοκληρωθούν. Η βελτιστοποίηση των ρυθμίσεων του Excel, η χρήση αποτελεσματικών τύπων και συναρτήσεων και η μείωση του μεγέθους των δεδομένων σας με το φιλτράρισμα ή τη σύνοψή τους μπορεί να συμβάλει στη βελτίωση των χρόνων απόκρισης για εργασίες ανάλυσης δεδομένων στο Excel.

Σχετικά με τις απαιτήσεις μνήμης RAM και την δύναμη του επεξεργαστή, παρατηρήθηκε πως υπάρχουν αρκετές απαιτήσεις σε μνήμη, γεγονός που δεν κάνει το excel την καλύτερη επιλογή για την εκτέλεση υπολογισμών και πόσο μάλιστα όταν πρόκειται για πιο σύνθετες ενέργειες. Αντιθέτως, στην επεξεργαστική δύναμη, οι απαιτήσεις δεν ήταν ίδιες και ανταποκρίθηκε σε πολύ καλό βαθμό χωρίς κάποια επιβάρυνση για τον υπολογιστή.

Ο απαιτούμενος αποθηκευτικός χώρος αποτελεί ένα σημαντικό μειονέκτημα για το συγκεκριμένο φύλλο εργασίας. Τα παραγόμενα αρχεία έχουν αρκετά μεγαλύτερο μέγεθος σε σύγκριση με γλώσσες προγραμματισμού (όπως για παράδειγμα η SQL). Τα xml αρχεία που προκύπτουν μπορούν να «κουβαλούν» τον ίδιο όγκο πληροφορίας, χρησιμοποιώντας πολύ λιγότερη μνήμη του υπολογιστή, γεγονός που τα καθιστά πολύ πιο αποδοτικά στην χρήση εξ' αιτίας της εξοικονόμησης χρόνου και χρόνου που προσφέρουν.

Για τα Ποιοτικά Χαρακτηριστικά:

Σχετικά με την φιλικότητα που προσφέρει προς τον χρήστη, αποτελεί ιδιαίτερα περιβάλλον και είναι απαραίτητο για την υιοθέτηση λογισμικού, την αποτελεσματικότητα, την ικανοποίηση των χρηστών και την εκπαίδευση. Δημιουργώντας ένα φιλικό προς τον χρήστη περιβάλλον, οι προγραμματιστές λογισμικού μπορούν να διασφαλίσουν ότι το λογισμικό τους είναι προσβάσιμο και αποτελεσματικό για χρήστες όλων των επιπέδων και των δεξιοτήτων.

Για την καλύτερη λειτουργία του Excel έχουν προστεθεί τα πρόσθετα του Excel (Excel Add-Ins) τα οποία είναι εργαλεία λογισμικού τρίτων που μπορούν να ενσωματωθούν στο Microsoft Excel για να βελτιώσουν τη λειτουργικότητά του. Μπορούν να παρέχουν πρόσθετες λειτουργίες, να αυτοματοποιήσουν επαναλαμβανόμενες εργασίες ή να βοηθήσουν στην ανάλυση δεδομένων. Ορισμένα

παραδείγματα πρόσθετων του Excel περιλαμβάνουν εργαλεία οπτικοποίησης δεδομένων, πρόσθετα ανάλυσης δεδομένων και εργαλεία οικονομικής μοντελοποίησης. Στο πλαίσιο της ανάλυσης δεδομένων, μερικά από τα πιο δημοφιλή πρόσθετα περιλαμβάνουν το Analysis Toolpak, το Solver και το MegaStat.

Τα πρόσθετα του Excel είναι κατάλληλα για μικρές εταιρείες και μικρά έργα λόγω της διαθεσιμότητας και του χαμηλού κόστους τους, ενώ τα SPSS, SAS και R λειτουργούν καλά για μεγάλα έργα και μεγάλες επιχειρήσεις λόγω της ικανότητάς τους να χειρίζονται αποτελεσματικά μεγάλα ποσά δεδομένων.

Επομένως, μεγαλύτερα σύνολα δεδομένων, όπως αυτά που συσσωρεύονται από τη Wal-Mart και το Facebook, απαιτούν λογισμικό υψηλότερης ισχύος, όπως SPSS, SAS ή R. Η διαφοροποίηση μεταξύ του ποιο από αυτά τα λογισμικά ταιριάζει καλύτερα στην ανάλυση αυτών των μεγαλύτερων συνόλων δεδομένων εξαρτάται από έναν αριθμό των παραγόντων, και κάθε στατιστικό πακέτο έχει τα δικά του δυνατά και αδύνατα σημεία. Ωστόσο σε κάθε περίπτωση πρέπει να έχει διερευνηθεί το κόστος, η διαδικασία εγκατάστασης, η απαραίτητη εκπαίδευση, τα ενσωματωμένα πακέτα, το επίπεδο επιθυμίας χειρισμού κώδικα, η υποστήριξη χρηστών, η σημασία των ελκυστικών γραφικών και πολλά άλλα ζητήματα με την ελπίδα να παρέχει σε επιχειρήσεις και πανεπιστήμια λεπτομέρειες που μπορούν να διευκολύνουν τις επιλογές λογισμικού τους. Συνεπώς με την τεράστια (και αυξανόμενη) εστίαση στα μεγάλα δεδομένα στη σημερινή κοινωνία, οι επιχειρήσεις, τα πανεπιστήμια και οι μελλοντικοί φοιτητές θα πρέπει να συμμετέχουν ενεργά. Τα μεγάλα δεδομένα είναι ένα μείγμα δομημένων και μη δομημένων δεδομένων και όχι απλώς ένας αριθμός εγγραφών. Οι δυνατότητες μεγάλων δεδομένων υπερβαίνουν τον μέγιστο αριθμό γραμμών και στηλών που μπορεί να χειριστεί ένα λογισμικό.

Από τη σκοπιά της ανταγωνιστικότητας και της πιθανής σύλληψης αξίας, όλες οι εταιρείες πρέπει να λαμβάνουν σοβαρά υπόψη τα μεγάλα δεδομένα. Στους περισσότερους κλάδους, οι εδραιωμένοι ανταγωνιστές και οι νεοεισερχόμενοι θα αξιοποιήσουν στρατηγικές που βασίζονται σε δεδομένα για να καινοτομήσουν, να ανταγωνιστούν και να αποσπάσουν αξία από πληροφορίες σε βάθος και σε πραγματικό χρόνο. (Manyika et al., 2011, σελ. 6)

Υπάρχουν διάφοροι τρόποι για την εκτέλεση ανάλυσης μεγάλων δεδομένων, αλλά η καλύτερη προσέγγιση θα εξαρτηθεί από τις συγκεκριμένες ανάγκες και απαιτήσεις της ανάλυσης. Ακολουθούν μερικές κοινές μέθοδοι:

1. Κατανεμημένος υπολογισμός: Χρησιμοποιήστε ένα κατανεμημένο υπολογιστικό πλαίσιο όπως το Hadoop ή το Spark για να επεξεργαστείτε και να αναλύσετε μεγάλους όγκους δεδομένων σε πολλαπλές μηχανές.
2. Υπηρεσίες που βασίζονται σε νέφος: Χρησιμοποιήστε υπηρεσίες που βασίζονται σε σύννεφο, όπως οι Υπηρεσίες Ιστού της Amazon ή το Microsoft Azure για την αποθήκευση και την επεξεργασία μεγάλων συνόλων δεδομένων. Αυτές οι υπηρεσίες προσφέρουν επεκτάσιμη, κατ' απαίτηση υποδομή για ανάλυση δεδομένων.
3. Εργαλεία οπτικοποίησης δεδομένων: Χρησιμοποιήστε εργαλεία οπτικοποίησης δεδομένων, όπως το Tableau ή το Power BI, για να βοηθήσετε στην ανάλυση και την

εξερεύνηση μεγάλων συνόλων δεδομένων και για να μεταδώσετε πληροφορίες στους ενδιαφερόμενους.

4. Μηχανική μάθηση και τεχνητή νοημοσύνη: Χρησιμοποιήστε αλγόριθμους μηχανικής μάθησης και τεχνητής νοημοσύνης για να προσδιορίσετε μοτίβα και τάσεις σε μεγάλα σύνολα δεδομένων και να κάνετε προβλέψεις και συστάσεις.

5. Προεπεξεργασία δεδομένων: Πριν από την ανάλυση, εκτελέστε βήματα προεπεξεργασίας δεδομένων, όπως καθαρισμό, μετασχηματισμό και συγκέντρωση των δεδομένων για να τα κάνετε πιο διαχειρίσιμα και σχετικά με την ανάλυση.

Συνολικά, η καλύτερη προσέγγιση για την ανάλυση μεγάλων δεδομένων είναι η χρήση ενός συνδυασμού αυτών των μεθόδων, προσαρμοσμένων στις συγκεκριμένες ανάγκες της ανάλυσης, λαμβάνοντας υπόψη όλους τους παράγοντες που αναφέρθηκαν παραπάνω (το κόστος, η διαδικασία εγκατάστασης, η απαραίτητη εκπαίδευση, τα ενσωματωμένα πακέτα, το επίπεδο επιθυμίας χειρισμού κώδικα, η υποστήριξη χρηστών, η σημασία των ελκυστικών γραφικών κα.).

ΒΙΒΛΙΟΓΡΑΦΙΑ

ΞΕΝΗ ΒΙΒΛΙΟΓΡΑΦΙΑ

Agrawal, G. (2021, November 24). A comprehensive guide to Microsoft Excel for Data Analysis. Retrieved from Analytics Vidhya website:

<https://www.analyticsvidhya.com/blog/2021/11/a-comprehensive-guide-on-microsoft-excel-for-data-analysis/>

Amadio, W., & Haywood, E. (2019). Data Analytics and the Cash Collections. Process: An adaptable Case Employing Excel and Tableau.

Article 29 Working Party (2013), Opinion 03/2013 on purpose limitation. Retrieved from:

https://ec.europa.eu/justice/article-29/documentation/opinion_recommendation/files/2013/wp203_en.pdf σελ.35

Ashraff, M. (2022, May). A Study of Big Data in Data Warehouse. *researchgate*.

Aydin, C., & Aydin, C. (2022, July 11). What Are the Benefits of Analysis ToolPak in Excel? Retrieved from projectcubicle. Retrieved from:

<https://www.projectcubicle.com/what-are-the-benefits-of-analysis-toolpak-in-excel/>

Birch, D., Lyford-Smith, D., & Guo, Y. (2018). The Future of Spreadsheets in the Big Data Era. ArXiv:1801.10231 [Cs]. Retrieved from:

<https://arxiv.org/abs/1801.10231>

Bock, A. (2016). A Literature Review of Spreadsheet Technology. Retrieved from:

https://pure.itu.dk/ws/files/81449279/ITU_TR_2016_199.pdf

Bose, B. (2019, June 22). How to Analyze Data in Excel: Simple Tips and Techniques.

Carey, P., & Berk, K. (2007). *Instructor's Solutions Manual for Data Analysis with Microsoft Excel*. Retrieved from:

https://testbank4textbook.com/pdf_samples/Solutions_Manual_for_Data_Analysis_with_Microsoft_Excel_Updated_for_Office_2007_3rd_Edition_by_Berk_sample_chapter.pdf

Chen, M., Mao, S., & Liu, B. (2014). *Big data: a survey*. (pp. 171–209).

Data Analysis. (2018). Analyzing Data Using Excel 1 Analyzing Data Using Excel Analyzing Data Using Excel Analyzing Data Using Excel. Retrieved from:

<https://www.vfu.bg/en/e-Learning/MS-Office--excel.pdf>

Data Analysis. (2022, July 28). [Updated] How to Add Data Analysis In Excel. Retrieved December 18, 2022. Retrieved from:

<https://sarasanalytics.com/blog/data-analysis-using-excel/>

Data Analysis with Excel. (2021). Data Analysis with Excel i. Retrieved from:
https://www.tutorialspoint.com/excel_data_analysis/excel_data_analysis_tutorial.pdf

En.wikipedia.org. (n.d.). *Big data*. [online] Retrieved from:
https://en.wikipedia.org/wiki/Big_data#cite_note-38 [Accessed 14 Sep. 2021].

Gartner IT Glossary. 2012. Big Data. Retrieved from:
<http://www.gartner.com/itglossary/big-data>.

Guerrero, H. =. (2019). Excel Data Analysis Modeling and Simulation.

Gupta, A. (2020, December 24). How to Add Data Analysis In Excel: The Best Guide [2022 Edition]. Retrieved from:

<https://www.simplilearn.com/tutorials/excel-tutorial/data-analysis-excel>

Manyika, J. et al., (2011). Big Data: The Next Frontier for Innovation, Competition, and Productivity: McKinsey Global Institute. Retrieved from:

https://bigdatawg.nist.gov/pdf/MGI_big_data_full_report.pdf

Meyer, D., & Avery, L. (2008). Excel as a Qualitative Data Analysis Tool.

Meyer, Z., & Avery, A. (2007). Organizing and analyzing qualitative data. In Teachers taking action: A comprehensive guide to teacher research.

Raymond A. Panko (2005). What We Know About Spreadsheets Errors. January 2005. Retrieved from:

https://www.researchgate.net/publication/228662532_What_We_Know_About_Spreadsheet_Errors

Richins, G., Stapleton, A., Stratopoulos, T. C., & Wong, C. (2017). Big Data Analytics: Opportunity or Threat for the Accounting Profession? *Journal of Information Systems*, 31(3), 63–79.

<https://doi.org/10.2308/isys-51805>

Shannon C., Weaver W., 1963. The Mathematical Theory of Communication. The University of Illinois: Urbana.

Shigarov, A., Khristyuk, V., Mikhailov, A., & Paramonov, V. (2019, May 1). Software Development for Rule-Based Spreadsheet Data Extraction and Transformation.

<https://doi.org/10.23919/MIPRO.2019.8756829>

Sodhar, I. H., Jalbani, A. H., Buller, A. H., & Naz, A. (2022, April). Data mining security for big data. *Security and Privacy Trends in Cloud Computing and Big Data*, σσ. 85-101.

Stucke M., Grunes A. (2016), *Big Data and Competition Policy*, Oxford University Press

Yang, W., Yan, Y., & Li, L. (2022, May). Data visualization and data analysis system of industrial Internet. *International Conference on Electronic Information Technology*

Technopedia. (2022). 5 Excel Data Analysis Tools to Know in 2022. Retrieved December 18, 2022. Retrieved from:

<https://www.techopedia.com/5-excel-data-analysis-tools-to-know-in-2022/2/34829>

Valli, K. (9 C.E., August 9). Importance of using excel in Data Analytics. Retrieved December 18, 2022. Retrieved from:

<https://www.linkedin.com/pulse/importance-using-excel-data-analytics-khadar-valli>

ΕΛΛΗΝΙΚΗ ΒΙΒΛΙΟΓΡΑΦΙΑ

Γιόγος Α., Μπούζας Α., (2022). Τεχνικές Ανάλυσης Επιχειρηματικών Δεδομένων Μεγάλου Όγκου με Υπολογιστικά Φύλλα. Μεσολόγγι.

Στεφανιδάκης, Μ., Ανδρόνικος, Θ., & Παπαδάκης, Ι. (2015). Ανοικτά συνδεδεμένα δεδομένα και εφαρμογές. Αθήνα: Κάλλιπος, Ανοικτές Ακαδημαϊκές Εκδόσεις.

Τσάκαλης, Π. (2020). Ανάλυση μεγάλων δεδομένων και δυναμική οπτικοποίηση αποτελεσμάτων. Πανεπιστήμιο Πειραιώς, Σχολή Τεχνολογιών Πληροφορικής και Επικοινωνιών. Τμήμα Ψηφιακών Συστημάτων. Πειραιάς: Τμήμα Ψηφιακών Συστημάτων-Πανεπιστήμιο Πειραιώς.