

Τίτλος Π.Ε. ΑΝΑΛΥΤΙΚΗ ΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ ΣΑΚΧΑΡΩΔΟΥΣ ΔΙΑΒΗΤΗ

Κωδικός Π.Ε. 19027

Όνοματεπώνυμο φοιτήτριας ΙΩΑΝΝΑ ΕΙΡΗΝΗ ΤΟΝΤΖΟΥ

Όνοματεπώνυμο εισηγητή ΔΗΜΗΤΡΙΟΣ Α. ΔΕΡΒΟΣ

Ημερομηνία ανάληψης Π.Ε. 2/04/2019

Ημερομηνία περάτωσης Π.Ε. 15/12/2020

Βεβαιώνω ότι είμαι η συγγραφέας αυτής της εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, έχω καταγράψει τις όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών, εικόνων και κειμένου, είτε αυτές αναφέρονται ακριβώς είτε παραφρασμένες. Επιπλέον, βεβαιώνω ότι αυτή η εργασία προετοιμάστηκε από εμένα προσωπικά, ειδικά ως πτυχιακή εργασία, στο Τμήμα Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων του ΔΙ.ΠΑ.Ε.

Η παρούσα εργασία αποτελεί πνευματική ιδιοκτησία της φοιτήτριας Τόντζου Ιωάννας Ειρήνης που την εκπόνησε. Στο πλαίσιο της πολιτικής ανοικτής πρόσβασης, ο συγγραφέας/δημιουργός εκχωρεί στο Διεθνές Πανεπιστήμιο της Ελλάδος άδεια χρήσης του δικαιώματος αναπαραγωγής, δανεισμού, παρουσίασης στο κοινό και ψηφιακής διάχυσης της εργασίας διεθνώς, σε ηλεκτρονική μορφή και σε οποιοδήποτε μέσο, για διδακτικούς και ερευνητικούς σκοπούς, άνευ ανταλλάγματος. Η ανοικτή πρόσβαση στο πλήρες κείμενο της εργασίας, δεν σημαίνει καθ' οιονδήποτε τρόπο παραχώρηση δικαιωμάτων διανοητικής ιδιοκτησίας του συγγραφέα/δημιουργού, ούτε επιτρέπει την αναπαραγωγή, αναδημοσίευση, αντιγραφή, πώληση, εμπορική χρήση, διανομή, έκδοση, μεταφόρτωση (downloading), ανάρτηση (uploading), μετάφραση, τροποποίηση με οποιονδήποτε τρόπο, τμηματικά ή περιληπτικά της εργασίας, χωρίς τη ρητή προηγούμενη έγγραφη συναίνεση του συγγραφέα/δημιουργού.

Η έγκριση της πτυχιακής εργασίας από το Τμήμα Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων του Διεθνούς Πανεπιστημίου της Ελλάδος, δεν υποδηλώνει απαραίτητα και αποδοχή των απόψεων του συγγραφέα, εκ μέρους του Τμήματος.

Στο τέλος μιας όμορφης συνεργασίας

Πρόλογος

Κατά τη διάρκεια των σπουδών μου στο Τμήμα Πληροφορικής του Αλεξάνδρειου Τεχνολογικού Εκπαιδευτικού Ιδρύματος (νυν: Τμήμα Πληροφορικής και Ηλεκτρονικής Μηχανικής του Διεθνούς Πανεπιστημίου της Ελλάδος), είχα την ευκαιρία να παρακολουθήσω τις ενότητες μαθημάτων με τίτλο «Εισαγωγή στην Ανάλυση Δεδομένων» και «Οργάνωση Δεδομένων και Εξόρυξη Πληροφορίας». Και τα δύο με παρακίνησαν στην περαιτέρω διερεύνηση της ανάλυσης δεδομένων μέσω του προγραμματισμού σε R. Τη σημασία της οπτικοποίησης στην ερμηνεία, την κατανόηση και την περαιτέρω αξιοποίηση της πληροφορίας που προκύπτει μέσα από τη διαδικασία της αναλυτικής επεξεργασίας δεδομένων.

Η παρούσα πτυχιακή εργασία αποτέλεσε ευκαιρία για ν' αναπτύξω γνώσεις και δεξιότητες στα εργαλεία και τις τεχνικές ανάλυσης δεδομένων που παρέχει το προγραμματιστικό περιβάλλον της R.

Κατά την εκπόνηση της πτυχιακής εργασίας, υπήρξε ουσιαστική συνεισφορά τεχνογνωσίας της Ιατρού Κυτταρολόγου κυρίας Χριστίνας Ζιώγα MD, MSc, MIAAC, στα ιατρικά θέματα της εργασίας και στην από ιατρική άποψη ερμηνεία και κατανόηση των αποτελεσμάτων.

Τα δεδομένα που χρησιμοποιήθηκαν προέρχονται από την εφαρμογή *forDiabetes* (<https://fordiabetes.app>), την οποία έχει αναπτύξει η εταιρεία Tesseract Multimedia (<https://www.tesseract.gr>) με χρηματοδότηση του προγράμματος ΕΣΠΑ. Η διάθεση των δεδομένων διέπεται από τους όρους συμφωνητικού εχεμύθειας με τίτλο «ΣΥΜΦΩΝΙΑ ΕΜΠΙΣΤΕΥΤΙΚΟΤΗΤΑΣ ΚΑΙ ΣΥΝΕΡΓΑΣΙΑΣ» η οποία έχει συνυπογραφεί με την εταιρεία Tesseract Multimedia για αποκλειστική χρήση των δεδομένων στο πλαίσιο της παρούσας πτυχιακής εργασίας και για την προστασία των προσωπικών δεδομένων των χρηστών της εφαρμογής.

Περίληψη

Η παρούσα πτυχιακή εργασία έχει ως κύριο στόχο τη διερεύνηση, επεξεργασία, οπτικοποίηση και ερμηνεία δεδομένων πασχόντων από Σακχαρώδη Διαβήτη, στο προγραμματιστικό περιβάλλον της R. Τα δεδομένα που χρησιμοποιήθηκαν είναι πραγματικά και προέρχονται από την εφαρμογή *forDiabetes* (<https://fordiabetes.app>) της εταιρείας Tessler Multimedia (<https://www.tessler.gr>).

Στην πορεία εκπόνησης της πτυχιακής εργασίας τα δεδομένα οργανώθηκαν ως προς τον τύπο του Σακχαρώδους Διαβήτη, την ηλικία και τη χώρα προέλευσης των ασθενών. Αξιολογήθηκε ο βαθμός ανταπόκρισης των χρηστών για κάθε ένα τύπο μέτρησης της εφαρμογής αλλά και συνδυαστικά, για τον εντοπισμό ικανού στατιστικού δείγματος χρηστών που χρησιμοποιούν με συστηματικό τρόπο την εφαρμογή για ένα ικανό συνεχόμενο χρονικό διάστημα και όχι περιστασιακά. Τα στατιστικά δείγματα που προέκυψαν αναλύθηκαν ως προς τα χαρακτηριστικά τους. Η οπτικοποίηση των αποτελεσμάτων βοήθησε στο να γίνει κατανοητή η ανάγκη για τη δημιουργία κινήτρων, ώστε οι χρήστες της εφαρμογής να εντάξουν την τελευταία στην καθημερινότητά τους και να τη χρησιμοποιούν με συστηματικό τρόπο.

Λαμβάνοντας υπόψη τα παραπάνω, διαμορφώθηκαν προτάσεις για την επέκταση της εφαρμογής *forDiabetes* με στόχο την πληρέστερη ενημέρωση του χρήστη, ώστε αυτός να έχει λόγους να αφιερώνει χρόνο στην καταχώρηση τιμών σε καθημερινή βάση. Διαμορφώθηκαν δύο τεκμηριωμένες προτάσεις: (α) η απεικόνιση σε ενιαίο γράφημα της μεταβολής των τριών βασικών μετρήσεων (γλυκόζη στο αίμα, ημερήσια καύση θερμίδων με την άσκηση, ημερήσια κατανάλωση υδατανθράκων) ώστε ο χρήστης να ερμηνεύει και να κατανοεί εύκολα την αμοιβαία εξάρτησή τους, και (β) η επέκταση της λειτουργικότητας της εφαρμογής, ώστε να καταχωρούνται με συστηματικό τρόπο και μετρήσεις που αφορούν στην ποιότητα του ύπνου του χρήστη. Υλοποιήθηκε προσομοίωση της πρότασης (α) στο περιβάλλον R/RStudio.

Κατά την εκπόνηση της πτυχιακής εργασίας και με την ολοκλήρωσή της, διεξήχθησαν δύο (2) συναντήσεις εργασίας με την εταιρεία Tessler Multimedia, όπου παρουσιάστηκαν τα αποτελέσματα και οι προτάσεις για επέκταση της εφαρμογής. Η συγκεκριμένη αλληλοεπίδραση με την εταιρεία Tessler Multimedia υπήρξε ιδιαίτερα εποικοδομητική και ανέδειξε στην πράξη τη σημασία της σύνδεσης της ακαδημαϊκής έρευνας με φορείς ανάπτυξης/παραγωγής.

ANALYTICAL PROCESSING OF DIABETES MELLITUS DATA

TONTZOY IOANNA EIRINI

Abstract

The aim of this final year thesis project has been to investigate, process, visualize and interpret data of Diabetes Mellitus patients in the R/RStudio programming environment. The data used are real and come from the *forDiabetes* application (<https://fordiabetes.app>) of Tessera Multimedia (<https://www.tessera.gr>).

Along the lines of pursuing the aforementioned goal, the data were organized according to the type of Diabetes, the age and the country of origin of the patients. The level of response of users was assessed separately for each type of measurement, as well as in combination, to identify a sufficient statistical sample of users who constantly use the application in a systematic way and not occasionally. The resulting statistical samples were analyzed as to their characteristics/features. The visualization of the results is expected to help the *forDiabetes* users realize the need to integrate the latter in their daily lives and use it in a systematic way.

In view of the above, an effort has been made to investigate ways to extend the functionality of the *forDiabetes* application in the direction of providing more complete information to the users who dedicate time to register their data on a daily basis. Two documented proposals were formed: (a) the representation in a single graph of the three basic measurements (blood glucose, daily calorie burning with exercise, daily carbohydrate consumption) so that the user can easily interpret and realize existing interdependencies between any two or all three measurements, (b) the extension of the functionality of the application so that measurements relating to the quality of sleep are also registered in a systematic way. Proposal (a) has been simulated in the R/RStudio environment utilizing the dataset provided.

During the elaboration of the thesis and upon its completion, two (2) working meetings were held with the company Tessera Multimedia. During these two meetings, the outcomes of the work herewith reported were presented to the *forDiabetes* development team as well as to the management of Tessera Multimedia. This interaction with the company has been very constructive and it has revealed in practice the importance of linking academic research to the needs of real development/production actors.

Ευχαριστίες

Με την περάτωση της παρούσας πτυχιακής εργασίας αισθάνομαι την ανάγκη να εκφράσω τις ειλικρινείς και θερμές ευχαριστίες μου σε όσους συνέβαλαν στην ολοκλήρωση αυτής της προσπάθειας:

Στον επιβλέποντα καθηγητή μου, κύριο Δημήτριο Δέρβο, για την πολύτιμη υποστήριξή του, τις παραγωγικές υποδείξεις του και το πολύ καλό κλίμα συνεργασίας που διαμόρφωσε.

Στην Ιατρό Κυτταρολόγο κυρία Χριστίνα Ζιώγα, χωρίς την συνεπίβλεψη και τις πολύτιμες ιατρικές υποδείξεις της οποίας δε θα ήταν δυνατή η ολοκλήρωση της πτυχιακής εργασίας.

Στην εταιρεία Tessera Multimedia, που διέθεσε τα δεδομένα της εφαρμογής *forDiabetes* και ήταν ανοιχτή σε συζητήσεις/συναντήσεις μέσα σε ένα άψογο κλίμα συνεργασίας.

Στην οικογένεια και τους φίλους μου, για τη συμπαράσταση και την υπομονή τους.

Περιεχόμενα

Πρόλογος.....	vii
Περίληψη	ix
Abstract.....	xi
Ευχαριστίες	xiii
Περιεχόμενα	xv
Κεφάλαιο 1ο: ΕΙΣΑΓΩΓΗ.....	1
1.1 Εισαγωγή	1
1.2 Βασικές Έννοιες.....	1
1.2.1 Δεδομένα και Πληροφορία.....	1
1.2.2 Σακχαρώδης Διαβήτης.....	3
1.3 Οργάνωση περιεχομένου	5
Κεφάλαιο 2ο: ΕΡΓΑΛΕΙΑ ΚΑΙ ΤΕΧΝΙΚΕΣ	7
2.1 Η γλώσσα προγραμματισμού σε R.....	7
2.1.1 Πακέτα (R packages)	7
2.1.2 Πλεονεκτήματα και Μειονεκτήματα	8
2.1.3 Εφαρμογές.....	9
2.1.4 RStudio.....	10
2.2 Ανάλυση Δεδομένων μέσω του προγραμματισμού R.....	11
2.3 Οπτικοποίηση Δεδομένων μέσω του προγραμματισμού σε R.....	11
2.3.1 Τύποι Γραφημάτων.....	12
2.3.2 Βιβλιοθήκη GGLOT2.....	15
Κεφάλαιο 3ο: ΤΑ ΔΕΔΟΜΕΝΑ.....	17
3.1 Σύνολο Δεδομένων – Προέλευση	17
3.2 Περιγραφή την Βάσης Δεδομένων	17
3.3 Μετασχηματισμοί Δεδομένων	19
3.3.1 Χαρακτηριστικά των χρηστών της εφαρμογής.....	19
3.3.2 Μετρήσεις της εφαρμογής.....	20
Κεφάλαιο 4ο: ΔΙΕΡΕΥΝΗΤΙΚΗ ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ: ΑΠΟΤΕΛΕΣΜΑΤΑ.....	23
4.1 Κατανομή των χρηστών βάσει των χαρακτηριστικών τους.....	23
4.1.1 Κατανομή αριθμού χρηστών ανά τύπο Σακχαρώδους Διαβήτη	23
4.1.2 Κατανομή αριθμού χρηστών ανά χώρα	24
4.1.3 Κατανομή αριθμού χρηστών ανά ήπειρο	25

4.1.4	Κατανομή αριθμού χρηστών ανά ηλικία.....	27
4.2	Αξιόπιστες περιπτώσεις χρηστών στις μετρήσεις της εφαρμογής	28
4.2.1	Κριτήρια Αξιολόγησης.....	29
4.2.2	Αλγόριθμος για την παραγωγή γραφημάτων	29
4.2.3	Απεικόνιση Γραφημάτων	30
4.2.4	Μέτρηση Activity	31
4.2.5	Μέτρηση Glucose	32
4.2.6	Μέτρηση Meal.....	33
4.2.7	Μέτρηση Medication	34
4.2.8	Μέτρηση Pressure.....	36
4.3	Ζευγάρια Μετρήσεων	37
4.3.1	Μετρήσεις Activity – Glucose.....	37
4.3.2	Μετρήσεις Meal – Glucose	38
4.3.3	Μετρήσεις Medication – Glucose.....	39
4.3.4	Μετρήσεις Pressure – Glucose	41
4.3.5	Μετρήσεις Activity – Glucose – Meal	42
Κεφάλαιο 5ο:	ΠΡΟΤΕΙΝΟΜΕΝΕΣ ΕΠΕΚΤΑΣΕΙΣ.....	45
5.1	Ενιαίο Γράφημα 3-σε-1	45
5.1.1	Κοινή κλίμακα μέτρησης	47
5.1.2	Οπτικοποίηση του Γραφήματος 3-σε-1	48
5.1.3	Χρηστική Διεπαφή.....	49
5.2	Εισαγωγή του ύπνου ως μέτρηση στην εφαρμογή	54
Κεφάλαιο 6ο:	ΕΠΙΛΟΓΟΣ	57
6.1	Σύνοψη.....	57
6.2	Σημασία της πτυχιακής εργασίας.....	57
Κεφάλαιο 7ο:	ΒΙΒΛΙΟΓΡΑΦΙΑ	59

Κεφάλαιο 1ο: ΕΙΣΑΓΩΓΗ

1.1 Εισαγωγή

Οι ραγδαίες εξελίξεις στον τομέα της Πληροφορικής και των Επικοινωνιών αποτέλεσαν τα θεμέλια για τη δημιουργία μιας νέας μορφής ηλεκτρονικής υγείας, την «κινητή υγεία» (m-Health). Ως κινητή υγεία ορίζεται η «πρακτική ιατρικής και δημόσιας υγείας που υποστηρίζεται από κινητές συσκευές, όπως κινητά τηλέφωνα, συσκευές παρακολούθησης ασθενών, προσωπικούς ψηφιακούς βοηθούς και άλλες ασύρματες συσκευές» [1].

Οι εφαρμογές υγείας που μπορούν να εγκατασταθούν στις έξυπνες συσκευές και οι φορητοί αισθητήρες επιτρέπουν τη συλλογή σημαντικών ιατρικών δεδομένων και δεδομένων που αφορούν τον τρόπο ζωής και την καθημερινή δραστηριότητα. Τα δεδομένα αυτά καταλαμβάνουν μεγάλο όγκο. Ανήκουν στην κατηγορία των big data και έλκουν ολοένα και περισσότερο το ενδιαφέρον της τεχνολογίας. Η σωστή ανάλυση και μελέτη αυτών μπορούν να συνεισφέρουν στην καλύτερη οργάνωση της βιομηχανίας της υγείας, στην παραγωγή προϊόντων και υπηρεσιών με επίκεντρο τις ανάγκες των ασθενών.

Η παρούσα πτυχιακή εργασία παρουσιάζει μία προσπάθεια ανάλυσης και αξιολόγησης των δεδομένων πραγματικών χρηστών, μέσα από μια υπάρχουσα εφαρμογή υγείας. Η εφαρμογή αυτή είναι η *forDiabetes* της εταιρείας Tesserà Multimedia, που αποσκοπεί στη διατήρηση ή βελτίωση υγιεινών συμπεριφορών μεμονωμένων διαβητικών ατόμων.

1.2 Βασικές Έννοιες

Στην παρούσα ενότητα παρατίθενται εισαγωγικές έννοιες της Πληροφορικής αλλά και της Ιατρικής αντίστοιχα, ώστε να μπορέσει ο αναγνώστης να αντιληφθεί τις ανάγκες που καλύπτει το παρόν πόνημα καθώς και τις προτάσεις που επιχειρεί να εισάγει στην εφαρμογή. Γίνεται αναφορά στον όρο δεδομένα και πληροφορία στον κόσμο την Πληροφορικής και αντίστοιχα στη νόσο του «σακχαρώδους διαβήτη» και στους διάφορους τύπους της.

1.2.1 Δεδομένα και Πληροφορία

Από την στιγμή που υπάρχουν πολλοί επιστημονικοί κλάδοι που μελετούν την πληροφορία και τις παραλλαγές της (δεδομένα, γνώση, σοφία) παρατηρούνται και διαφορές στον τρόπο με το οποίο αναφέρονται στις αντίστοιχες θεμελιώδεις έννοιες. Σύμφωνα με τη μελέτη [2], παρατίθενται ο ορισμός για τα δεδομένα και την πληροφορία.

Για τον ορισμό των δεδομένων, το [2] βασίστηκε στον ορισμό του McCrank για τα δεδομένα και τα γεγονότα.

Ο McCrank (2002, σελ 627) [3] ορίζει τα δεδομένα ως την πρώτη ύλη της πληροφορίας. Αποτελούν τις μικρότερες μονάδες κάποιας πληροφορίας οι οποίες ανακαλύπτονται, συγκεντρώνονται, βγαίνουν συμπεράσματα και επεξεργάζονται. Φυσικά δεν είναι γεγονότα αλλά ούτε μπορούν να παρουσιαστούν από μόνα τους.

Για να ξεπεραστούν οι διαφορές που υπάρχουν ως προς τον όρο των δεδομένων στους διαφορετικούς επιστημονικούς κλάδους ο McCrank όρισε τα γεγονότα. Ως γεγονότα ορίζει τις ενέργειες και τις συμπεριφορές που προέρχονται μέσα από κάτι ήδη γνωστό, κάποιο πρόβλημα, το οποίο έχει

πραγματική υπόσταση και σχετίζεται με την αντικειμενική πραγματικότητα. Με την υπόθεση ότι τα γεγονότα είναι σταθερά και πραγματικά, μπορούν να καταστούν προφανή.

Τα γεγονότα μπορούν μόνο να καταγράφονται και στη συνέχεια να επεξεργάζονται. Η φάση στην οποία εμπλέκονται τα δεδομένα είναι όταν τα γεγονότα πρέπει να παρουσιαστούν σωστά στο κατάλληλο μοντέλο. Ο ορισμός για τα δεδομένα είναι ότι αυτά παρουσιάζουν τα γεγονότα του πραγματικού κόσμου [2].

Για παράδειγμα μπορούν να συγκροτήσουν το αποτέλεσμα των μετρήσεων που διενεργήθηκαν γύρω από ένα φαινόμενο του πραγματικού κόσμου (οι τιμές της θερμοκρασίας μιας δεδομένης γεωγραφικής περιοχής σε μια συγκεκριμένη χρονική περίοδο).

Τα δεδομένα διατίθενται σε διάφορους τύπους. Το χαρακτηριστικό που διαχωρίζει τις δυο βασικές κατηγορίες δεδομένων είναι η αλληλεξάρτηση μεταξύ των πεδίων που τα συγκροτούν.

Από την μία υπάρχουν δεδομένα μεταβλητών (πεδία) μεταξύ των οποίων δεν υπάρχει αμοιβαία εξάρτηση. Βάσει των τιμών των δεδομένων τους, διακρίνονται δύο βασικές κατηγορίες μεταβλητών/πεδίων:

- Μεταβλητές οι οποίες παίρνουν αριθμητικού τύπου τιμές, π.χ. 15.2, 37,19, κλπ.
- Μεταβλητές οι οποίες καταχωρούν κατηγορικού τύπου δεδομένα, π.χ. «μήλο», «πορτοκάλι», κλπ.

Τα δεδομένα (αριθμητικού ή κατηγορικού τύπου) διακρίνονται περαιτέρω σε [4]:

- Ονομαστικά (nominal): προσδιορίζουν τις μεταβλητές χωρίς κάποια αριθμητική τιμή. Είναι «ετικέτες» στις μεταβλητές. Για παράδειγμα η μεταβλητή για το φύλο του ανθρώπου, περιέχει τις ετικέτες άνδρας και γυναίκα.
- Τακτικά (ordinal): οι μεταβλητές προσδιορίζονται από ταξινομημένες κατηγορίες και οι διαφορές μεταξύ των κατηγοριών δεν είναι γνωστές. Για παράδειγμα σε μια κλίμακα για την διάθεση ενός ατόμου, δεν είναι γνωστές οι διαφορές μεταξύ του «ΟΚ», «Χαρούμενος» και του «Πολύ Χαρούμενος».
- Διαστήματος (interval): οι μεταβλητές παίρνουν αριθμητικές τιμές στις οποίες είναι γνωστό τόσο η σειρά όσο και οι ακριβείς διαφορές μεταξύ των τιμών. Χαρακτηριστικό παράδειγμα είναι οι βαθμοί Κελσίου της θερμοκρασίας, οι 70 βαθμοί είναι μικρότεροι από τους 80 βαθμούς και η διαφορά τους είναι 10 βαθμοί, το ίδιο ισχύει και για τους 50 με 60 βαθμούς. Το πρόβλημα με τα διαστημικά τύπου δεδομένα είναι ότι δεν υπάρχει «πραγματικό μηδέν», όπως στο παράδειγμα το μηδέν δεν σημαίνει απουσία τιμής αλλά 0 βαθμούς Κελσίου.
- Κλασματικά (ratio): παίρνουν αριθμητικές τιμές που έχουν τις ίδιες ιδιότητες με τα διαστημικά δεδομένα. Η μόνη διαφορά είναι στα κλασματικά υπάρχει το «πραγματικό μηδέν». Χαρακτηριστικά παραδείγματα είναι το ύψος και το βάρος.

Τέλος, στο [5] αναφέρονται επιπλέον κατηγορίες δεδομένων, όπως τα:

- Δεδομένα με Μεικτά Γνωρίσματα, ορισμένα πεδία λαμβάνουν αριθμητικές και άλλα κατηγορηματικές τιμές.
- Δυαδικά Δεδομένα, το πεδίο μπορεί να λάβει μόνο μία από τις δυο αριθμητικές ή κατηγορηματικές επιτρεπόμενες τιμές.

Στον αντίποδα υπάρχουν δεδομένα μεταβλητών στα οποία υπάρχει έμμεση ή άμεση σχέση εξάρτησης μεταξύ των μεταβλητών/πεδίων. Διακρίνονται οι ακόλουθοι τύποι δεδομένων :

- Δεδομένα χρονοσειρών, αποτελούνται από δυο πεδία. Το ένα είναι η χρονική στιγμή και το άλλο μια αριθμητική τιμή. Στο παράδειγμα με την τιμή της θερμοκρασίας μια δεδομένη χρονική στιγμή.
- Διακριτές Ακολουθίες και Συμβολοσειρές, έχουν ανάλογη δομή με τα δεδομένα χρονοσειρών. Το ένα πεδίο είναι η χρονική στιγμή ή ένα αριθμητικό αναγνώρισης θέσης και το άλλο μια κατηγορηματική τιμή.
- Χωρικά Δεδομένα, σε αυτό τον τύπο δεδομένων οι μετρήσεις προκύπτουν με βάση την τοποθεσία. Τα υπόλοιπα γνωρίσματα μπορούν να είναι αριθμητικά ή κατηγορηματικά. Για παράδειγμα, οι μετεωρολόγοι συλλέγουν τις θερμοκρασίες της επιφάνεια της θάλασσας για να προβλέψουν τους τυφώνες. Σε αυτή την περίπτωση το ένα πεδίο είναι οι συντεταγμένες της επιφάνειας και στο άλλο οι θερμοκρασίες.

Η άμεση ανάθεση ενός ορισμού για την πληροφορία είναι πολύ δύσκολη [2]. Αρχικά ο Matclup (1980) όρισε τη γνώση ως περιεχόμενο και την πληροφορία ως την διαδικασία εξεργασίας. Στην προσπάθειά του να δικαιολογήσει τον συλλογισμό του περιορίστηκε στην δραστηριότητα των επιστημόνων και των ερευνητών και δεν έλαβε υπόψη την επεξεργασία που έγινε από τους βιβλιοθηκονόμους και διαχειριστές αρχείων. Αργότερα, ο Matclup και ο Mansfield παρατήρησαν ότι η πληροφορία είναι κάτι διαφορετικό ανάλογα με την οπτική γωνία που εξετάζεται. Πιο συγκεκριμένα σημαίνει κάτι διαφορετικό για αυτούς που την εξετάζουν ως προς τα χαρακτηριστικά, τις ιδιότητες, τα στοιχεία, τις τεχνικές, τις λειτουργίες, τις διαστάσεις και τις συνδέσεις της.

Παρόλα αυτά, η πληροφορία φαίνεται να σχετίζεται με την επικοινωνία και την ερμηνεία των δεδομένων. Η έννοια της πληροφορίας γίνεται καλύτερα κατανοητή από τους φορείς επιρροής που διαθέτουν πολλές και εύκολα προσπελάσιμες πληροφορίες για το περιβάλλον.

Λαμβάνοντας αυτά υπόψη, η πληροφορία ορίζεται ως η επιτυχής ερμηνεία των δεδομένων, που στόχο έχει ν' αυξήσει το όφελος, το κέρδος ή και την ευχαρίστηση εφόσον πραγματοποιείται μέσω κάποιας πνευματικής δραστηριότητας [2].

1.2.2 Σακχαρώδης Διαβήτης

Ο σακχαρώδης διαβήτης (ΣΔ) είναι μια κοινή και χρόνια πάθηση, που χαρακτηρίζεται από υψηλά επίπεδα γλυκόζης (σακχάρου) στο αίμα. Εμφανίζεται είτε όταν το πάγκρεας δεν παράγει ικανοποιητικά ινσουλίνη, είτε λόγω ελαττωμένης ευαισθησίας των κυττάρων στη δράση της ινσουλίνης.

Σε φυσιολογικές συνθήκες, η ινσουλίνη βοηθά τη γλυκόζη να εισέλθει στα κύτταρα, εξασφαλίζοντάς τους ενέργεια. Στην περίπτωση που κάποιος πάσχει από ΣΔ η γλυκόζη δεν καταφέρνει να εισέλθει στα κύτταρα, με αποτέλεσμα να αυξάνεται το ποσοστό της στο αίμα.

Ο ΣΔ θεωρείται μεταβολικό νόσημα. Μεταβολισμός είναι το σύνολο των βιοχημικών αντιδράσεων που πραγματοποιούνται στα κύτταρα των ζωντανών οργανισμών και περιλαμβάνει όλες εκείνες τις βιοχημικές διαδικασίες, που εμπλέκονται στην παραγωγή και την απελευθέρωση της ενέργειας. Όταν τρώμε, ο οργανισμός μας διασπά και μετασχηματίζει τις τροφές σε μία πολύ απλή μορφή, που ονομάζεται γλυκόζη. Η γλυκόζη τροφοδοτεί κάθε κύτταρο του σώματος με την ενέργεια που αυτό χρειάζεται και μεταφέρεται, μέσω του αίματος και των αγγείων, στα κύτταρα όπου χρησιμοποιείται ή στους χώρους αποθήκευσης. Η γλυκόζη, όμως, δεν μπορεί μόνη της να εισέλθει μέσα στα κύτταρα. Χρειάζεται την ινσουλίνη, μια ορμόνη που απελευθερώνεται από το πάγκρεας και λειτουργεί σαν κλειδί, επιτρέποντας την είσοδό της στα κύτταρα. Όταν η γλυκόζη εισέρχεται στα κύτταρα, μειώνεται

το ποσοστό της στο αίμα. Στην περίπτωση που κάποιος πάσχει από ΣΔ, η γλυκόζη δεν μπορεί να εισέλθει στα κύτταρα, με αποτέλεσμα ν' αυξάνεται η συγκέντρωσή της στο αίμα. Με την πάροδο του χρόνου, η υψηλή συγκέντρωση γλυκόζης στο αίμα μπορεί να προκαλέσει προβλήματα υγείας και να οδηγήσει σε σοβαρές επιπλοκές, όπως νεφρική ανεπάρκεια και καρδιαγγειακά νοσήματα. Γι' αυτό είναι απαραίτητο τα άτομα με ΣΔ να ελέγχουν διαρκώς τα επίπεδα γλυκόζης στο αίμα τους.

Υπάρχουν πέντε (5) τύποι Σακχαρώδους Διαβήτη [6] [7] [8] [9]:

Ο Σακχαρώδης Διαβήτης τύπου 1, παλαιότερα γνωστός ως ινσουλινοεξαρτώμενος ή νεανικός διαβήτης, είναι μια ασθένεια που συνήθως εμφανίζεται σε παιδική ή εφηβική ηλικία. Ο ίδιος ο οργανισμός, μέσω αντισωμάτων που παράγει, επιτίθεται και καταστρέφει τα β-κύτταρα του παγκρέατος, που με τη σειρά τους παράγουν ινσουλίνη. Ως αποτέλεσμα είναι να υπάρχει ελάχιστη ή μηδενική παραγωγή ινσουλίνης.

Η εξέλιξη του ΣΔ τύπου 1 είναι συνήθως σταδιακή, ενώ τα συμπτώματα τις περισσότερες φορές εκδηλώνονται αφού ο οργανισμός έχει ήδη ξεκινήσει την καταστροφή των β-κύτταρων του παγκρέατος. Ο ΣΔ τύπου 1 δεν είναι κληρονομικός, αλλά για την εκδήλωσή του υπάρχει γενετική προδιάθεση, που δεν είναι δυνατό να ανατραπεί ούτε από το άτομο που θα εκδηλώσει τελικά τη νόσο, αλλά ούτε -προς το παρόν- από την επιστήμη. Επίσης, περιβαλλοντικοί παράγοντες, όπως κάποιοι ιοί, το άγχος, τα αλλεργιογόνα των τροφών, μπορεί να συμβάλλουν εξίσου στην εμφάνισή του. Έχει ακόμα βρεθεί ότι ανάμεσα σε διαφορετικές εθνικότητες ή φυλές, ο ΣΔ τύπου 1 έχει διαφορετικά ποσοστά εμφάνισης. Τα άτομα που πάσχουν από ΣΔ τύπου 1, θα πρέπει να λαμβάνουν εξωγενώς ινσουλίνη για να επιβιώσουν. Ο σχεδιασμός και η τήρηση ενός πλάνου υγιεινών γευμάτων και προγράμματος σωματικής δραστηριότητας, καθώς και ο έλεγχος των επιπέδων σακχάρου στο αίμα, είναι οι ακρογωνιαίοι λίθοι για τη διατήρηση της υγείας των ασθενών με ΣΔ.

Ο Σακχαρώδης Διαβήτης τύπου 2, παλαιότερα γνωστός ως μη ινσουλινοεξαρτώμενος ή διαβήτης των ενηλίκων, αποτελεί την πιο κοινή μορφή ΣΔ. Τον τύπο αυτό φέρουν περίπου 80% των ατόμων που πάσχουν από ΣΔ. Στο ΣΔ τύπου 2, το πάγκρεας δεν παράγει αρκετή ινσουλίνη ή το σώμα δεν τη χρησιμοποιεί αποτελεσματικά. Στον τύπο αυτό υπάρχει κληρονομική προδιάθεση, δηλαδή εκδηλώνεται πιο συχνά σε άτομα με ιστορικό διαβήτη στο οικογενειακό τους περιβάλλον. Επίσης, μπορεί να είναι το αποτέλεσμα ενός τρόπου ζωής που παρουσιάζει το συνδυασμό υπερβολικού βάρους και έλλειψης σωματικής δραστηριότητας. Ένας αυξανόμενος αριθμός παιδιών και εφήβων έχει διαγνώσει με ΣΔ τύπου 2 τα τελευταία χρόνια. Όπως και στον ΣΔ τύπου 1, οι τροφές, η σωματική δραστηριότητα και η φαρμακευτική αγωγή, επηρεάζουν τα επίπεδα γλυκόζης αίματος. Ο έλεγχος των επιπέδων γλυκόζης του αίματος βοηθάει στη διαχείριση της πάθησης.

Ο Σακχαρώδης Διαβήτης της Κύησης είναι μια πάθηση κατά τη διάρκεια της εγκυμοσύνης, όπου εμφανίζονται υψηλά επίπεδα γλυκόζης στο αίμα. Μπορεί να γίνει διάγνωση μέσω εξέτασης, όταν η γυναίκα βρίσκεται μεταξύ της 24ης και της 28ης εβδομάδας της κύησης. Έχει ποσοστό εκδήλωσης περίπου 7% του συνόλου των κυήσεων και διατρέχει κινδύνους τόσο η μητέρα όσο και το έμβryo. Οι περισσότερες γυναίκες μπορούν να διαχειριστούν τον συγκεκριμένο τύπο, με αλλαγές στον τρόπο ζωής τους, σε συνδυασμό με ένα πλάνο υγιεινών γευμάτων. Ακόμη, υπάρχουν περιπτώσεις γυναικών που χρειάζεται να λάβουν και ινσουλίνη. Επιπλέον, στις περισσότερες γυναίκες με ΣΔ της Κύησης, τα επίπεδα της γλυκόζης του αίματος επανέρχονται στα φυσιολογικά, μετά τη γέννηση του βρέφους. Ωστόσο, γυναίκες που εκδήλωσαν ΣΔ της Κύησης είναι πιο πιθανό να αναπτύξουν ΣΔ τύπου 2.

Η πάθηση, κατά την οποία τα επίπεδα γλυκόζης του αίματος είναι υψηλότερα από τα φυσιολογικά, αλλά όχι αρκετά υψηλά ώστε να διαγνωστεί διαβήτης, ονομάζεται προδιαβήτης. Στα άτομα που έχουν

διαγνωστεί με προδιαβήτη, υπάρχει αυξημένος κίνδυνος ανάπτυξης ΣΔ τύπου 2 ως μετέπειτα στάδιο. Ο προδιαβήτης σχετίζεται με την παχυσαρκία -ιδιαίτερα γύρω από την κοιλιακή χώρα ή τα εσωτερικά όργανα-, με μη φυσιολογική ποσότητα χοληστερόλης (δυσλιπιδαιμία) και με την υψηλή πίεση αίματος. Για την πρόληψη ή την καθυστέρηση της εμφάνισης ΣΔ τύπου 2, συνιστάται τα άτομα με προδιαβήτη να προσπαθήσουν να μειώσουν τη μάζα του σώματος τους κατά 7% και να αυξήσουν τη σωματική τους άσκηση σε τουλάχιστον 150 λεπτά μέτριας έντασης σωματικής δραστηριότητας εβδομαδιαίως. Φαρμακευτική αγωγή μπορεί, επίσης, να συνταγογραφηθεί, προκειμένου να βοηθήσει στη βελτίωση της ανοχής στη γλυκόζη.

Ο Σακχαρώδης Διαβήτης τύπου LADA, δηλαδή ο Λανθάνων Αυτοάνοσος Διαβήτης σε Ενήλικες, είναι μια μορφή που παρουσιάζει το μηχανισμό του ΣΔ τύπου 1, αλλά με πιο αργούς ρυθμούς καταστροφής των β-κυττάρων του παγκρέατος, ενώ έχει πολλά κοινά χαρακτηριστικά κλινικά γνωρίσματα με τον ΣΔ τύπου 2. Σε αντίθεση με τον ΣΔ τύπου 1, ο LADA συχνά δεν απαιτεί τη χορήγηση ινσουλίνης για αρκετούς μήνες έως και χρόνια μετά τη διάγνωση. Είναι πολύ πιθανό, όμως, να απαιτήσει θεραπεία με ινσουλίνη κάποια στιγμή στο μέλλον. Ο έλεγχος των επιπέδων σακχάρου στο αίμα είναι εξίσου πολύ σημαντικός στον LADA. Η διαίτα με περιορισμένο γλυκαιμικό δείκτη αποτελεί την κυρία οδό στην αντιμετώπιση του, ειδικά στις νέες περιπτώσεις ασθενών.

1.3 Οργάνωση περιεχομένου

Στο κεφάλαιο 2 παρουσιάζονται τα εργαλεία και οι τεχνικές του προγραμματισμού σε R που χρησιμοποιήθηκαν κατά την συγγραφή του κώδικα για την παράγωγή των στατιστικών στοιχείων και των γραφικών παραστάσεων. Γίνεται αναφορά και στο θεωρητικό κομμάτι της R ώστε να μπορέσει ο αναγνώστης να κατανοήσει με ευκολία τα αποτελέσματα της διερευνητικής ανάλυσης που ακολουθεί.

Στο κεφάλαιο 3 γίνεται η παρουσίαση των πρωτογενών δεδομένων. Παρουσιάζεται η βάση των δεδομένων που χρησιμοποιήθηκε με το διάγραμμα ER. Ακολουθεί ο μετασχηματισμός των δεδομένων και η τελική μορφή που χρησιμοποιήθηκε στην είσοδο στο στάδιο της ανάλυσης για την παραγωγή στατιστικών στοιχείων και γραφημάτων.

Το κεφάλαιο 4 περιλαμβάνει τα αποτελέσματα της διερευνητικής ανάλυσης των δεδομένων. Συγκεκριμένα περιέχει τα γραφήματα που δημιουργήθηκαν συνοδευόμενα από το αλγόριθμο παραγωγής τους και την ερμηνεία τους.

Στο κεφάλαιο 5 παρατίθενται οι δύο (2) προτάσεις επέκτασης της εφαρμογής, που στόχο έχουν την κινητοποίηση των χρηστών να καταχωρούν πιο συχνά δεδομένα στην εφαρμογή. Περιλαμβάνει το γράφημα 3-σε-1 της μεταβολής τριών διαφορετικών μετρήσεων και την εισαγωγή του ύπνου ως νέο τύπο μέτρησης στην εφαρμογή.

Στο κεφάλαιο 6 συνοψίζονται τα αποτελέσματα της πτυχιακής εργασίας και γίνεται αναφορά στη σημασία και τα οφέλη της.

Κεφάλαιο 2ο: ΕΡΓΑΛΕΙΑ ΚΑΙ ΤΕΧΝΙΚΕΣ

Σε αυτό το κεφάλαιο γίνεται παρουσίαση των εργαλείων που αξιοποιήθηκαν στη συγγραφή του κώδικα για την παραγωγή στατιστικών στοιχείων και στη συνέχεια τη γραφική απεικόνισή τους. Γίνεται αναφορά στη γλώσσα προγραμματισμού R. Πιο συγκεκριμένα παρουσιάζονται διαδοχικά τα πακέτα που χρησιμοποιήθηκαν αλλά και το περιβάλλον ανάπτυξης RStudio. Επιπλέον περιγράφονται τα βήματα της ανάλυσης των δεδομένων μέσω του προγραμματισμού σε R. Δίνεται έμφαση στο βήμα της οπτικοποίησης καθώς παρουσιάζονται οι πιο γνωστοί τύποι R γραφημάτων και το πακέτο ggplot2 που χρησιμοποιήθηκε.

2.1 Η γλώσσα προγραμματισμού σε R

Ο κώδικας της πτυχιακής εργασίας υλοποιήθηκε σε γλώσσα προγραμματισμού R. Η R διαθέτει έναν εκτενή κατάλογο στατιστικών και γραφικών μεθόδων. Χάρη στις δυνατότητες που προσφέρει χρησιμοποιείται ευρέως στη στατιστική έρευνα, ανάλυση δεδομένων, δημιουργία γραφημάτων και υλοποίηση αλγόριθμων Μηχανικής Μάθησης.

Ανήκει στη λίστα των ελεύθερων λογισμικών, ανοιχτού κώδικα, που παρέχεται υπό τους όρους άδειας GNU (General Public License). Μπορεί να μεταλειτουργήσει και να τρέξει σε όλες τις γνωστές πλατφόρμες λειτουργικών UNIX, Windows, MacOS και Linux [10].

Ακόμη, δημοσκοπήσεις και έρευνες εξόρυξης δεδομένων δείχνουν σημαντική αύξηση της δημοτικότητας της. Μάλιστα, από το Σεπτέμβριο του 2020, η R κατατάσσεται στην 9η θέση του δείκτη TIOBE, ένα μέτρο δημοτικότητας των γλωσσών προγραμματισμού [11].

2.1.1 Πακέτα (R packages)

Ο πυρήνας της R είναι αρκετά ισχυρός και συναρπαστικός, όμως τα πακέτα είναι αυτά που κάνουν την R πραγματικά το απαραίτητο εργαλείο στον τομέα των δεδομένων και της μηχανικής μάθησης.

Τα πακέτα της R είναι συλλογές μεθόδων (functions) και συνόλων δεδομένων (datasets). Αποθηκεύονται κάτω από έναν κατάλογο που ονομάζεται "βιβλιοθήκη". Για να μπορεί να γίνει χρήση των μεθόδων ενός πακέτου θα πρέπει πρώτα να εγκατασταθεί η αντίστοιχη βιβλιοθήκη.

```
# install package
install.packages("dplyr")

# load package
library(dplyr)
```

Εικόνα 2.1.1.1 Εγκατάσταση βιβλιοθήκης και φόρτωση των πακέτων της.

Το CRAN αποτελεί το παγκόσμιο χώρο αποθήκευσης πακέτων. Συντηρείται από το ίδρυμα R. Τα πακέτα που ανεβαίνουν εδώ πρέπει να περάσουν πολλές δοκιμές για να διασφαλιστεί ότι ακολουθούνται οι πολιτικές της CRAN. Μάλιστα υπάρχουν διαθέσιμα περισσότερα από 16.000 πακέτα που επεκτείνουν τις δυνατότητες της R [12].

Η R έρχεται με ένα τυπικό σύνολο πακέτων. Με τη βοήθεια της εντολής `search()`, μπορεί κανείς να βρει όλη τη λίστα των διαθέσιμων πακέτων που είναι εγκατεστημένα στο σύστημα. Ακολουθεί ο

κατάλογος με τα πιο γνωστά πακέτα αλλά και αυτά που χρησιμοποιήθηκαν στην πτυχιακή. Η λίστα έχει ως εξής:

- **TIDYR**: όπως δηλώνει και το όνομα της, χρησιμοποιείται για τον μετασχηματισμό των δεδομένων σε μια πιο κατάλληλη μορφή για επεξεργασία στο προγραμματιστικό περιβάλλον της R [13].
- **GGPLOT2**: το πακέτο αυτό περιέχει μεθόδους δημιουργίας γραφικών παραστάσεων. Φημίζεται για τα κομψά και ποιοτικά γραφήματα, που το κάνουν να ξεχωρίζει από τα άλλα πακέτα οπτικοποίησης [14].
- **DPLYR**: το πακέτο αυτό χρησιμοποιείτε για τον χειρισμό των δεδομένων. Είναι χτισμένο κυρίως γύρω από αυτές τις 5 εξαιρετικές και εύχρηστες λειτουργίες: `select()`, `filter()`, `arrange()`, `mutate()`, `summarize()`. Εστιάζει στη διαχείριση των `data frames` [15].
- **SCALES**: πακέτο που παρέχει εργαλεία για την επεξεργασία των υπομνημάτων και των ετικετών για τους άξονες των γραφικών παραστάσεων [16].
- **READXL**: πακέτο για γρήγορη εισαγωγή δεδομένων από ένα excel αρχείο στην R [17].
- **READR**: πακέτο που παρέχει ένα γρήγορο και φιλικό τρόπο ανάγνωσης δομημένων δεδομένων (όπως `csv`, `tsv` και `fwf`) [18].
- **SHINY**: πακέτο που βοηθάει στην ανάπτυξη διαδραστικών και αισθητικά ευχάριστων εφαρμογών ιστού χρησιμοποιώντας την R [19].
- **MARKDOWN**: πακέτο για την παρουσίαση των δεδομένων. Επιτρέπει τη δημιουργία αναφορών σε `HTML`, `Word` και άλλες μορφές [20].

2.1.2 Πλεονεκτήματα και Μειονεκτήματα

Η R είναι μια από της πιο δημοφιλής γλώσσες προγραμματισμού για στατιστική μοντελοποίηση και ανάλυση δεδομένων. Αλλά όπως και κάθε άλλη γλώσσα προγραμματισμού, η R έχει το δικό της σύνολο πλεονεκτημάτων και περιορισμών.

Σε αυτή την ενότητα παρουσιάζονται τα προτερήματα και τα μειονεκτήματα του προγραμματιστικού περιβάλλοντος τη R. Η R είναι μια συνεχώς εξελισσόμενη γλώσσα. Αυτό σημαίνει ότι πολλά από τα μειονεκτήματα σταδιακά θα εξαφανιστούν με τις μελλοντικές ενημερώσεις της R [21].

Το βασικό της πλεονέκτημα, αποτελεί το γεγονός ότι πρόκειται για μια γλώσσα προγραμματισμού ανοιχτού κώδικα. Αυτό σημαίνει ότι μπορεί γίνει χρήση της χωρίς κάποια ειδική άδεια ή χρέωση. Εκτελείται πολύ εύκολα σε όλα τα λειτουργικά συστήματα, όπως `Linux`, `Mac` και `Windows`

Ακόμη, παρέχει υποστήριξη για τη διαχείριση δεδομένων. Πακέτα όπως τα `dplyr`, `readr` είναι ικανά να μετασχηματίσουν τα ακατέργαστα δεδομένα σε κατάλληλη μορφή για να επεξεργαστούν από την R. Επιπλέον οι δημοφιλείς βιβλιοθήκες της όπως η `ggplot2` και `plotly` συμβάλουν στην παραγωγή οπτικά ελκυστικών γραφημάτων που κάνουν την R να ξεχωρίζει από τις άλλες γλώσσες προγραμματισμού.

Ένα ακόμη θετικό χαρακτηριστικό της είναι το ότι μπορεί να συνδυαστεί με πολλές άλλες γλώσσες προγραμματισμού όπως `C`, `C++`, `Java` και `Python`.

Με τα πακέτα όπως το `Shiny` και το `Markdown`, η αναφορά των αποτελεσμάτων μιας ανάλυσης είναι εξαιρετικά εύκολη. Μπορεί κανείς να συντάξει αναφορές με τα δεδομένα, τις γραφικές παραστάσεις και τα κομμάτια κώδικα R που είναι ενσωματωμένα σε αυτά. Ακόμη, παρέχει τη δυνατότητα να

δημιουργηθούν διαδραστικές εφαρμογές ιστού που επιτρέπουν στον χρήστη να παίζει με τα δεδομένα και τα αποτελέσματα.

Από την άλλη πλευρά αρχικό μειονέκτημα της R είναι η «αδύναμη προέλευση» της. Πηγή της είναι, μια παλαιότερη γλώσσα προγραμματισμού, η "S". Αυτό σημαίνει ότι το βασικό πακέτο της δεν υποστηρίζει τη δημιουργία δυναμικών ή τρισδιάστατων γραφικών. Τη λύση προσφέρουν τα κοινά πακέτα ggplot και plotly που αναφέρονται παραπάνω.

Το βασικό αρνητικό στοιχείο της, είναι η μεγάλη κατανάλωση μνήμης RAM. Επίσης, σε σχέση με τους ανταγωνιστές της (Python, Matlab), θεωρείται πιο αργή γλώσσα προς τον χρόνο εκτέλεσης των εντολών της.

2.1.3 Εφαρμογές

Οι εφαρμογές της R δεν περιορίζονται σε ένα μόνο τομέα. Η εφαρμογή της στο τραπεζικό τομέα, το ηλεκτρονικό εμπόριο και την οικονομία αποτελούν μόνο μερικά παραδείγματα ανάμεσα στους τομείς που έχει υιοθετηθεί. Ακολουθεί εκτενής κατάλογος με τις περιπτώσεις χρήσης του προγραμματισμού της R στον πραγματικό κόσμο.

Στην ακαδημαϊκή κοινότητα η R χρησιμοποιείται τόσο από στατιστικούς όσο και σπουδαστές για την εκτέλεση στατιστικών υπολογισμών και αναλύσεων. Περιπτώσεις χρήσης της σε αυτό το τομέα αποτελούν:

- Το Cornell University συνιστά στους ερευνητές και σπουδαστές την R για όλες τις έρευνες που αφορούν τη στατιστική πληροφορική.
- Το UCLA χρησιμοποιεί την R για να διδάξει στατιστική και ανάλυση δεδομένων στους σπουδαστές του.

Στον τομέα της πληροφορικής χρησιμοποιούν την R για να δημιουργήσουν εργαλεία και υπηρεσίες διαχείρισης δεδομένων. Μερικά παραδείγματα χρήσης της είναι:

- Το Mozilla για την απεικόνιση της δραστηριότητας του προγράμματος περιήγησης Firefox.
- Η Microsoft εντός του πλαισίου Azure Machine Learning.
- Η Google υπολογίζει τι κέρδος έχει παραχθεί σε σύγκριση με τις δαπάνες για διαφήμιση.

Επιπλέον, η R στον τομέα της οικονομίας παρέχει ένα προηγμένο σύνολο λειτουργιών για όλες τις οικονομικές εργασίες και υπολογισμούς. Μέσω του προγραμματιστικού περιβάλλοντος της R πραγματοποιείται πολύ εύκολα η μοντελοποίηση του χρηματιστηρίου, η ανάλυση χρηματοοικονομικών δεδομένων και η εκτίμηση κινδύνων. Η Lloyds of London και Bajaj Allianz Insurance αποτελούν παραδείγματα σε αυτό τον τομέα.

Αδιαμφισβήτητα οι σύγχρονες αγορές έχουν μετακινηθεί στο Διαδίκτυο. Η βιομηχανία ηλεκτρονικού εμπορίου κάνει χρήση της R στη στρατηγική μάρκετινγκ, στη στοχευμένη διαφήμιση και στη μοντελοποίηση των πωλήσεων. Για παράδειγμα η Amazon χρησιμοποιεί την R για να βελτιώσει τις προτάσεις μεταξύ των προϊόντων.

Οι εταιρίες μέσω κοινωνικής δικτύωσης χρησιμοποιούν την R για την ανάλυση της συμπεριφοράς και των συναισθημάτων των χρηστών τους. Έχουν τη δυνατότητα να βελτιώνουν τις προτάσεις τους βάσει του ιστορικού των αναρτήσεων των χρηστών τους. Με ανάλογο τρόπο προσαρμόζονται και οι διαφημίσεις που εμφανίζονται. Παραδείγματα αποτελούν το Twitter αλλά και το Facebook μέχρι το 2017.

Οι τραπεζικές εταιρείες χρησιμοποιούν την R για τη μοντελοποίηση του πιστωτικού κινδύνου και άλλες μορφές ανάλυσης κινδύνου. Επίσης, χρησιμοποιείται για την ανίχνευση απάτης, μοντελοποίηση ενυπόθηκων δανείων, αξιολόγηση πελατών και πολλά άλλα. Εκτός από τα στατιστικά στοιχεία, οι τράπεζες χρησιμοποιούν την R για την οπτικοποίηση των δεδομένων τους. Οι λειτουργίες αυτές αξιοποιούνται από την ANZ Bank και την Τράπεζα της Αμερικής (BOA).

Η γενετική, η βιοπληροφορική και η επιδημιολογία είναι ορισμένοι τομείς της βιομηχανίας υγειονομικής περίθαλψης που χρησιμοποιούν την R σε μεγάλο βαθμό. Χρησιμοποιείται για την ανάλυση και την πρόβλεψη της εξάπλωσης διαφόρων ασθενειών και στην ανάλυση δεδομένων για την ασφάλεια των φαρμάκων. Το πακέτο BioConductor της R παρέχει δυνατότητες ανάλυσης γονιδιωματικών δεδομένων. Η Merck & Co ανήκει στη λίστα με τις εταιρείες που κάνουν χρήση της R γλώσσας προγραμματισμού.

Τέλος διάφορες κατασκευαστικές εταιρείες χρησιμοποιούν την R για να ενισχύσουν τις στρατηγικές μάρκετινγκ τους. Αναλύουν τα σχόλια των πελατών τους για τη βελτίωση των προϊόντων τους. Μια επιπλέον περίπτωση χρήσης της ανάλυσης δεδομένων της R είναι για την πρόβλεψη της ζήτησης και των τάσεων της αγοράς. Κατασκευαστικές εταιρείες που κάνουν χρήση της R είναι:

- Η Ford Motor Company για την παραγωγή στατιστικών στοιχείων σχετικά με την επιχειρηματική στρατηγική που θα ακολουθήσουν. Βοηθάει να βελτιώσουν τα μελλοντικά τους σχέδια μέσα από την ανάλυση της αντίδρασης των πελατών της σχετικά με τα προϊόντα τους
- Η John Deere για την πρόβλεψη της ζήτησης των προϊόντων και των ανταλλακτικών της.

2.1.4 RStudio

Το RStudio [22] είναι ένα ολοκληρωμένο περιβάλλον ανάπτυξης (IDE) για την R. Είναι διαθέσιμο υπό την GNU Affero General Public License έκδοση 3. Το AGPL v3 επιτρέπει την ελεύθερη κοινοποίηση του κώδικα. Ακόμη, είναι γραμμένο σε γλώσσες προγραμματισμού Java, C ++ και JavaScript.

Επιπλέον, διατίθεται σε δύο εκδόσεις. Το RStudio Desktop στο οποίο τα προγράμματα εκτελούνται ως τοπικές εφαρμογές. Ενώ το RStudio Server εκτελείται σε έναν απομακρυσμένο διακομιστή και επιτρέπει την πρόσβαση στο RStudio χρησιμοποιώντας ένα πρόγραμμα περιήγησης ιστού.

Και οι δυο εκδοχές διατίθενται με εμπορικές και ανοιχτού κώδικα εκδόσεις. Το GUI του RStudio βασίζεται στο C ++ και Qt framework. Μερικές από τις δυνατότητες του RStudio είναι οι εξής [23]:

- Το αποκλειστικό IDE για την R. Το RStudio επιτρέπει την επισήμανση της σύνταξης και την εκτέλεση του προγράμματος απευθείας από τον επεξεργαστή προέλευσης.
- Βελτιώνει τη ροή εργασίας. Το RStudio είναι ικανό να διαχειρίζεται πολλούς καταλόγους (directories) με τη βοήθεια των προγραμμάτων της αποτελεσματικά
- Ικανότητα εντοπισμού σφαλμάτων και σύνταξης. Το RStudio διαγιγνώσκει και διορθώνει τα σφάλματα σε λιγότερο χρόνο. Όχι μόνο αυτό, περιλαμβάνει εκτενή κατάλογο εργαλείων ανάπτυξης πακέτων συγγραφής με Sweave και R Markdown.

Ακόμη, μπορεί να κάνει πλήρη ανάλυση δεδομένων χρησιμοποιώντας την R και άλλες γλώσσες προγραμματισμού. Το RStudio Desktop είναι διαθέσιμο για τα λειτουργικά συστήματα Windows,

macOS και Linux. Τα RStudio Server και Server Pro εκτελούνται σε Debian, Ubuntu, Linux, CentOS, openSUSE και SLES.

2.2 Ανάλυση Δεδομένων μέσω του προγραμματισμού R

Σε αυτή την ενότητα παρουσιάζονται όλα τα απαραίτητα βήματα για την ανάλυση δεδομένων μέσω του προγραμματισμού της R. Η όλη διαδικασία περιλαμβάνει πέντε βήματα, από την εισαγωγή των δεδομένων μέχρι την επικοινωνία των αποτελεσμάτων.

Το πρώτο βήμα είναι η εισαγωγή. Σε αυτό το στάδιο γίνεται η εισαγωγή των δεδομένων στο προγραμματιστικό περιβάλλον της R. Αυτό σημαίνει ότι τα δεδομένα που είναι αποθηκευμένα σε αρχεία, βάσεις δεδομένων ή σε οποιαδήποτε άλλη μορφή μετατρέπονται σε R data frames για να ξεκινήσει η ανάλυση τους.

Το δεύτερο βήμα είναι ο μετασχηματισμός. Στόχος σε αυτό το στάδιο είναι η μετατροπή των δεδομένων του data frame σε κατάλληλη μορφή για να επεξεργαστούν στην R. Κάθε στήλη του data frame είναι μια μεταβλητή και κάθε σειρά μια παρατήρηση της. Ακόμη, περιλαμβάνει το φιλτράρισμα των δεδομένων του data frame για τον περιορισμό των παρατηρήσεων και τη δημιουργία νέων μεταβλητών που είναι συναρτήσεις των ήδη υπαρχών μεταβλητών. Το βήμα αυτό ολοκληρώνεται με την παραγωγή στατιστικών στοιχείων των παρατηρήσεων.

Το τρίτο βήμα είναι η οπτικοποίηση. Αυτό το στάδιο βοηθάει στην κατανόηση των δεδομένων μέσω της αναπαράστασης τους σε γραφήματα. Μέσω της ανάγνωσης των γραφημάτων είναι πιο εύκολη η εύρεση μοτίβων ή και εξαιρέσεων στα δεδομένα. Η οπτικοποίηση των δεδομένων αποτελεί τον πιο γρήγορο και αποτελεσματικότερο τρόπο μετάδοσης της πληροφορίας.

Το τέταρτο βήμα αποτελείται από τα μοντέλα. Τα μοντέλα είναι συμπληρωματικά εργαλεία οπτικοποίησης. Είναι συνήθως μαθηματικά εργαλεία, που δίνουν απαντήσεις σε ερωτήσεις που σχετίζονται με τις παρατηρήσεις των δεδομένων.

Το πέμπτο βήμα είναι η επικοινωνία. Το τελευταίο στάδιο της ανάλυσης του προγραμματισμού σε R εστιάζει στην επικοινωνία και κοινοποίηση των αποτελεσμάτων. Η R παρέχει εργαλεία για την παραγωγή καλά σχεδιασμένων και ποιοτικών αναφορών.

2.3 Οπτικοποίηση Δεδομένων μέσω του προγραμματισμού σε R

Ένα από τα βασικά χαρακτηριστικά της R που την κάνουν να ξεχωρίζει από τις άλλες γλώσσες προγραμματισμού είναι η δυνατότητα παραγωγής ποιοτικών γραφημάτων για την παρουσίαση των δεδομένων.

Τα πακέτα της R επιτρέπουν τη δημιουργία οπτικά ελκυστικών γραφημάτων με μόνο λίγες γραμμές κώδικα. Η οπτικοποίηση των δεδομένων αποτελεί αποτελεσματική τεχνική απόκτησης πληροφοριών σχετικά με τα δεδομένα.

Ο ανθρώπινος εγκέφαλος επεξεργάζεται πιο αποτελεσματικά τις πληροφορίες που αναπαριστώνται σε γραφήματα. Οι τεχνικές οπτικοποίησης διευκολύνουν τον εντοπισμό κρυμμένων μοτίβων σε δεδομένα που σε διαφορετική περίπτωση θα αγνοούνταν. Επιπλέον, στην περίπτωση επεξεργασία μεγάλων συνόλων δεδομένων βοηθούν στην πιο γρήγορη κατανόηση της βασικής πληροφορίας που περιέχουν.

Κάθε στατιστικό γράφημα αποτελείται από κάποια βασικά στοιχεία, γνωστά ως γραμματική των γραφημάτων [24].

```

ggplot(continent_diab, aes(x= Continent, group= Type)) +
+   geom_bar(aes(y = ..prop.., fill = factor(..x..)), stat="count") +
+   geom_text(aes( label = scales::percent(..prop..),
+                 y= ..prop.. ), stat= "count", vjust = -.5) +
+   labs(y = "Percent", fill="Continent") +
+   facet_wrap(~Type) +
+   scale_y_continuous(labels = scales::percent)
+   theme_light()

```

Εικόνα 2.3.1 Απόσπασμα κώδικα που περιέχει όλα τα στοιχεία της γραμματικής των γραφημάτων.

Σύμφωνα με την Εικόνα 2.3.1, που αντιστοιχεί στο κώδικα για την παραγωγή ενός γραφήματος, η γραμματική των γραφημάτων περιέχει τα εξής στοιχεία:

- **Data:** είναι το σύνολο των δεδομένων και περιέχει τις μεταβλητές που θα σχεδιαστούν στο γράφημα.
- **Aesthetic:** Αφορά την κλίμακα πάνω στην οποία θα γίνει η αναπαράσταση των δεδομένα. Ουσιαστικά προσδιορίζει το σημείο της κάθε παρατήρησης στο γράφημα, αντιστοιχίζοντας τις μεταβλητές του συνόλου δεδομένων με τις x,y μεταβλητές του γραφήματος.
- **Geometries:** Αναφέρεται στο τρόπο με τα οποίο θα σχεδιαστούν τα δεδομένα, μπορεί να είναι σημεία, γραμμές και πολλά άλλα.
- **Facet:** Το faceting χωρίζει τα δεδομένα σε υποομάδες και σχεδιάζει υπογραφήματα για κάθε ομάδα.
- **Statistics:** αναφέρεται στην αναπαράσταση στατιστικών πληροφοριών σχετικά με τα δεδομένα, όπως μέσος όρος και διακύμανση, για να βοηθήσουμε στην κατανόηση των δεδομένων.
- **Coordinates:** αναφέρεται στο χώρο στον οποίο σχεδιάζονται τα γραφήματα. Τα πιο γνωστά γραφήματα χρησιμοποιούν το καρτεσιανό σύστημα συντεταγμένων.
- **Themes:** επιτρέπουν τις αλλαγές στα γραφήματα που δεν αφορούν τα δεδομένα αλλά τη γραμματοσειρά, την παλέτα χρωμάτων και άλλα σχεδιαστικά στοιχεία.

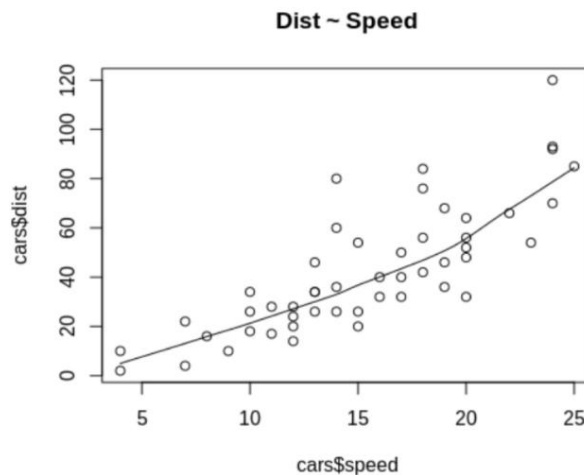
2.3.1 Τύποι Γραφημάτων

Ο προγραμματισμός με R συμβάλλει στην εκμάθηση των τεχνικών της οπτικοποίησης των δεδομένων προσφέροντας ένα σύνολο ενσωματωμένων λειτουργιών και τις συναρτήσεις που οργανώνονται σε βιβλιοθήκες. Τα γραφήματα που παράγονται ανήκουν σε έναν από τους παρακάτω βασικούς τύπους παρουσίασης:

- **Σύγκριση,** χρησιμοποιείται για την απεικόνιση των διαφορών που υπάρχουν σε ένα σύνολο όπως το ραβδόγραμμα
- **Σύνθεση,** χρησιμοποιείται για την απεικόνιση των στοιχείων που συγκροτούν το σύνολο όπως τα γραφήματα πίτας
- **Κατανομή,** χρησιμοποιείται για την απεικόνιση του πως διαμοιράζονται τα στοιχεία του συνόλου βάσει κάποιου χαρακτηριστικού τους όπως τα διαγράμματα διασποράς
- **Σχέση,** χρησιμοποιείται για την απεικόνιση του τρόπου με τον οποίο δύο ή περισσότερα στοιχεία συνδέονται μεταξύ τους, όπως το διάγραμμα γραμμών.

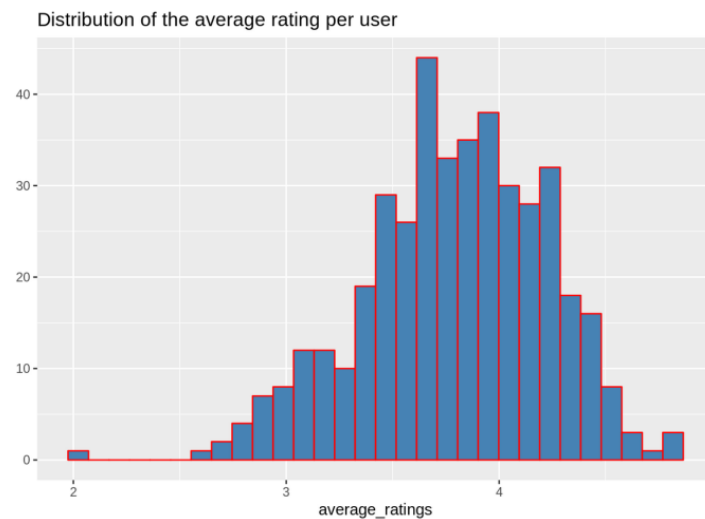
Παρακάτω παρατίθενται τα γραφήματα που χρησιμοποιούνται περισσότερο στην οπτικοποίηση δεδομένων συνοδευόμενα με τις περιπτώσεις χρήσης τους.

Αρχικά υπάρχει το **διάγραμμα διασποράς** (scatter plot) μέσα από το οποίο εξετάζεται η σχέση μεταξύ δυο συνεχόμενων μεταβλητών. Η Εικόνα 2.3.1.1 αντιστοιχεί στο διάγραμμα διασποράς και αναπαριστά την απόσταση που διανύει ένα αυτοκίνητο με την ταχύτητα που αυτό έχει.



Εικόνα 2.3.1.1 Διάγραμμα Διασποράς.

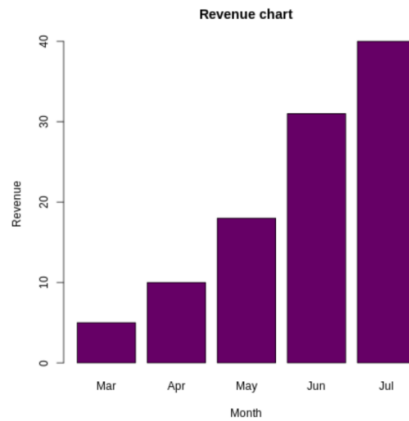
Ακολουθεί το **ιστογράμμο** (histogram). Αυτός ο τύπος γραφήματος χρησιμοποιείται για την απεικόνιση μιας συνεχούς μεταβλητής. Διαχωρίζει τα δεδομένα σε μικρότερα κομμάτια/ομάδες (bins) και απεικονίζει την κατανομή της συχνότητας αυτών των κομματιών.



Εικόνα 2.3.1.2 Ιστογράμμο.

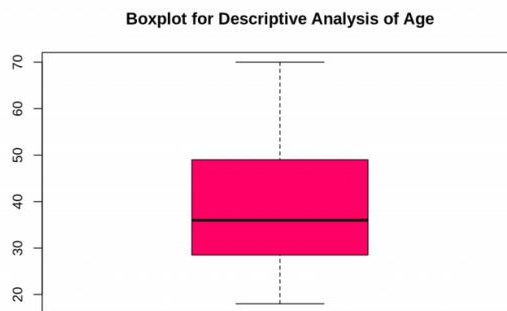
Η Εικόνα 2.3.1.2 αντιστοιχεί σε ένα ιστογράμμο και αναπαριστά την κατανομή των χρηστών βάσει της μέσης βαθμολογίας τους.

Το **διάγραμμα ράβδων** (bar chart) χρησιμοποιείται για τη γραφική απεικόνιση της κατανομής των τιμών μεταβλητής. Η Εικόνα 2.3.1.3 είναι ένα διάγραμμα ράβδων, δείχνει και συγκρίνει τα έσοδα για κάθε μήνα.



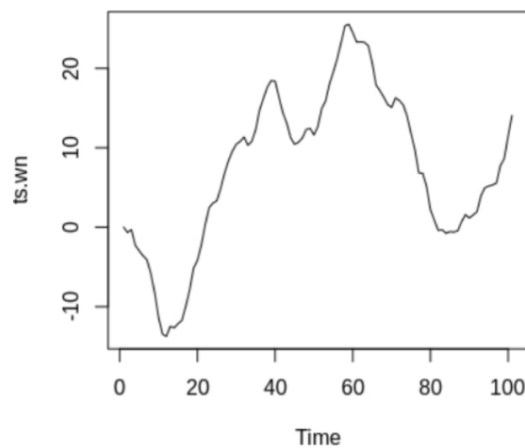
Εικόνα 2.3.1.3 Διάγραμμα Ράβδων.

Τα **box plot** γραφήματα χρησιμοποιούνται για την απεικόνιση αριθμητικού τύπου μεταβλητές. Επίσης, χρησιμοποιείται για την οπτικοποίηση της διασποράς των τιμών και για τον εντοπισμό ακραίων τιμών. Υπολογίζουν και οπτικοποιούν πέντε στατιστικά μεγέθη: το ελάχιστο, το ποσοστημόριο (quantile) 25%, το ποσοστημόριο (quantile) 75%, και την μέγιστη τιμή.



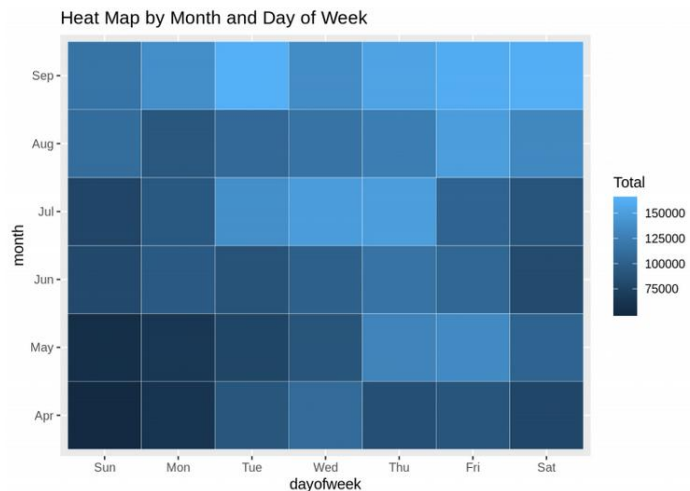
Εικόνα 2.3.1.4 Γράφημα box plot.

Το **διάγραμμα περιοχής** (area chart) χρησιμοποιείται για την ανάδειξη της μεταβολής μιας μεταβλητής ή ενός συνόλου δεδομένων. Επίσης, χρησιμοποιείται για την γραφική απεικόνιση χρονοσειρών. Εναλλακτικά χρησιμοποιείται για την ανάλυση των επιμέρους συνιστωσών συνεχών μεταβλητών.



Εικόνα 2.3.1.5 Διάγραμμα Περιοχής.

Το γράφημα τύπου **heat map** κωδικοποιεί και απεικονίζει την αμοιβαία εξάρτηση δύο ή περισσότερων μεταβλητών με την ένταση των χρωμάτων του. Ισοδύναμα, εμπλουτίζει την πληροφορία ενός δυσδιάστατου γραφήματος με την εισαγωγή μιας τρίτης διάστασης, την ένταση του χρώματος.



Εικόνα 2.3.1.6 Heat map γράφημα.

Το γράφημα στην εικόνα 2.3.1.6 είναι τύπου heat map γράφημα και αναπαριστά τη σχέση μεταξύ της μεταβλητής month στον y άξονα, τη μεταβλητή dayofweek στο x άξονα και τις πωλήσεις/έσοδα που αντιστοιχεί στα χρώματα.

2.3.2 Βιβλιοθήκη GGPLOT2

Στο CRAN, τον παγκόσμιο χώρο αποθήκευσης των R πακέτων, υπάρχουν διαθέσιμα περισσότερα από 16.000 πακέτα. Αυτό σημαίνει ότι υπάρχουν πακέτα που καλύπτουν σχεδόν κάθε ανάγκη οπτικοποίησης των δεδομένων. Στην παρούσα πτυχιακή εργασία έγινε χρήση του γραφικού πακέτου ggplot2 για την οπτικοποίηση των αποτελεσμάτων που προέκυψαν από τη διερευνητική ανάλυση των δεδομένων.

Το πακέτο ggplot2 αποτελεί ένα ολοκληρωμένο σύστημα σχεδίασης γραφημάτων. Αναπτύχθηκε από τον Hadley Wickham ως εφαρμογή της γραμματικής των γραφικών. Χρησιμοποιείται κυρίως για τη δημιουργία γραφημάτων επαγγελματικής εμφάνισης. Για τη σχεδίαση τους απαιτείται λίγος χρόνος και ελάχιστος κώδικας.

Τα δεδομένα που εισάγονται έχουν τη μορφή των data frames. Αυτό σημαίνει ότι κάθε σειρά είναι μια παρατήρηση και κάθε στήλη είναι μια μεταβλητή. Το ggplot2 είναι ένα σχετικά ευέλικτο και ισχυρό πακέτο γραφικών. Μπορεί να κάνει πολλά πράγματα, αλλά δεν μπορεί να δημιουργήσει τρισδιάστατα γραφήματα.

Η προγραμματιστική διεπαφή του ggplot2, επιτρέπει τον προσδιορισμό των μεταβλητών που θα σχεδιαστούν στο γράφημα, τον τρόπο εμφάνισης και τις οπτικές ιδιότητες. Επομένως χρειάζονται ελάχιστες αλλαγές στο κώδικα για να αλλάξει ο τρόπος εμφάνισης, για παράδειγμα από ραβδόγραμμα σε διάγραμμα διασποράς.

Τα ggplot2 γραφήματα δημιουργούνται βήμα-βήμα, προσθέτοντας κάθε φορά νέα στάδια/επίπεδα επεξεργασίας. Αυτή η δόμηση προσθέτει ευελιξία και προσαρμοστικότητα στη δημιουργία γραφημάτων.

R

```
ggplot(data = <DATA>, mapping = aes(<MAPPINGS>)) + <GEOM_FUNCTION>()
```

Εικόνα 2.3.2.1 Βασική σύνταξη ggplot2 για τη δημιουργία γραφήματος.

Σύμφωνα με την Εικόνα 2.3.2.1 που αντιστοιχεί στη συνάρτηση ggplot2, για τη δημιουργία του γραφήματος χρειάζονται τα εξής βήματα:

- Μέσω του ορίσματος data προσδιορίζεται το data frame που περιέχει τις μεταβλητές που θα αναπαρασταθούν στο γράφημα.
- Στο όρισμα mapping καλείται η συνάρτηση aes() για να καθοριστούν οι μεταβλητές και ο τρόπος παρουσίασης τους στο γράφημα. Για παράδειγμα ορίζεται η μεταβλητή που αναπαρασταθεί στον άξονα x και y, ή τα χαρακτηριστικά του όπως μέγεθος και χρώμα.
- Για την προσθήκη ενός νέου πεδίου για τη παραγωγή του γραφήματος, χρησιμοποιείται το (+) σύμβολο.
- Προσθήκη “geoms”. Από τη συνάρτηση geom() επιλέγεται το γράφημα αναπαράστασης των μεταβλητών (σημεία, γραμμές, ράβδοι). Το ggplot2 προσφέρει πολλά διαφορετικά γεωμετρικά στοιχεία. Τα πιο γνωστά είναι:
 - geom_point(), για το διάγραμμα διασποράς και διάγραμμα σημείων.
 - geom_boxplot(), για το ραβδόγραμμα.
 - geom_line(), για το διάγραμμα γραμμών.

Κεφάλαιο 3ο: ΤΑ ΔΕΔΟΜΕΝΑ

Σε αυτό το κεφάλαιο γίνεται η παρουσίαση των πρωτογενών δεδομένων. Δίνεται η προέλευση και η περιγραφή της βάσης των δεδομένων. Ακολουθεί ο μετασχηματισμός των δεδομένων και η τελική μορφή που χρησιμοποιήθηκε στην είσοδο στο στάδιο της ανάλυσης για την παραγωγή στατιστικών στοιχείων και γραφημάτων.

3.1 Σύνολο Δεδομένων – Προέλευση

Τα δεδομένα που χρησιμοποιήθηκαν προέρχονται από την εφαρμογή *forDiabetes* της εταιρείας Tesseract (<https://fordiabetes.app/author/tesseract/>). Η εφαρμογή αυτή χρησιμοποιείται από άτομα που πάσχουν από Σακχαρώδη Διαβήτη για να καταγράφουν και να ελέγχουν τους παράγοντες που επηρεάζουν τις τιμές της γλυκόζης στο αίμα στην καθημερινή του ζωής.

Η εφαρμογή βάσης δεδομένων *forDiabetes* λειτουργεί σε έναν XAMPP Server. Στο πλαίσιο της παρούσας πτυχιακής εργασίας, έγιναν διαθέσιμα από την εταιρεία και χρησιμοποιήθηκαν δεδομένα μέχρι τον Σεπτέμβριο του 2020. Διασφαλίστηκε η απόκρυψη των ευαίσθητων προσωπικών στοιχείων των χρηστών. Για την ανάκτηση των πρωτογενών δεδομένων έχει συνταχθεί κώδικας SQL.

Για τις ανάγκες της πτυχιακής εργασίας συγκεντρώθηκαν και αναλύθηκαν τα δεδομένα της εφαρμογής που αφορούν:

- Τα χαρακτηριστικά των χρηστών.
- Τη σωματική άσκηση των χρηστών.
- Τα επίπεδα γλυκόζης στο αίμα των χρηστών.
- Τα γεύματα που κατανάλωσαν οι χρήστες.
- Τα φάρμακα που χορηγήθηκαν στους χρήστες.
- Την αρτηριακή πίεση των χρηστών.

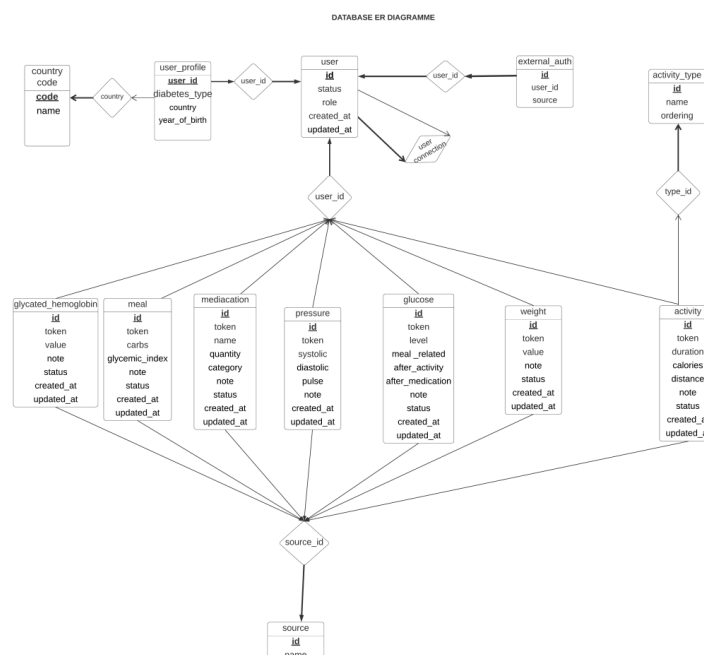
3.2 Περιγραφή την Βάσης Δεδομένων

Για την παρουσίαση των δεδομένων που είναι διαθέσιμα για την εργασία ακολουθεί περιήγηση στη βάση δεδομένων *forDiabetes* μέσα από το διάγραμμα ER της Εικόνας 3.2.1. Το διάγραμμα ER αποτελεί την οπτική αναπαράσταση διαφορετικών οντοτήτων μέσα σε ένα σύστημα και πώς σχετίζονται μεταξύ τους. Στο διάγραμμα οι κόμβοι αναπαριστούν τα σύνολα οντοτήτων και περιέχουν τα γνωρίσματα από τα οποία αποτελείται κάθε οντότητα. Οι ρόμβοι αναπαριστούν το σύνολο των σχέσεων που υπάρχουν μεταξύ των οντοτήτων. Τα βέλη ενώνουν τις οντότητες και τις σχέσεις που υπάρχουν μεταξύ τους.

Όπως απεικονίζεται στην Εικόνα 3.2.1, η βάση δεδομένων αποτελείται από δεκατρείς (13) πίνακες:

- Ο πίνακας **user** καταχωρεί δεδομένα που αφορούν στον λογαριασμό του χρήστη, πιο συγκεκριμένα: αν είναι διαπιστευμένος ή επισκέπτης. Επιπλέον χαρακτηρίζει αν ο χρήστης είναι ασθενής ή γιατρός. Τέλος περιέχει και την ημερομηνία δημιουργίας λογαριασμού του χρήστη.

- Ο πίνακας **user_profile** περιέχει πιο προσωπικά δεδομένα του χρήστη. Περιέχει τον τύπο σακχαρώδους διαβήτη, τη χώρα αλλά και το έτος γέννησης που δήλωσε όταν δημιούργησε τον λογαριασμό του στην εφαρμογή.
- Ο πίνακας **country_code** αντιστοιχεί στον διεθνή κωδικό μιας χώρας (πεδίο country) του πίνακα user_profile με το πλήρες όνομά της.
- Ο πίνακας **external_auth** περιγράφει αν ο χρήστης δημιούργησε το λογαριασμό του μέσω κάποιου άλλου λογαριασμού κοινωνικής πλατφόρμας παραδείγματος χάρι facebook, google.
- Ο πίνακας μέτρησης **activity** περιέχει εγγραφές που αφορούν στη σωματική άσκηση του χρήστη.
- Ο πίνακας **activity_type** προσδιορίζει τον κωδικό δραστηριότητας (πεδίο type_id) του πίνακα activity σε ποια δραστηριότητα αντιστοιχεί για παράδειγμα 1=περπάτημα, 2=κολύμβηση.
- Ο πίνακας μέτρησης **weight** περιέχει εγγραφές που αφορούν στο βάρος του χρήστη.
- Ο πίνακας μέτρησης **glucose** περιέχει εγγραφές για τα επίπεδα γλυκόζης του χρήστη.
- Ο πίνακας μέτρησης **pressure** περιέχει εγγραφές για την αρτηριακή πίεση του χρήστη.
- Ο πίνακας μέτρησης **medication** περιέχει εγγραφές που αφορούν στη φαρμακευτική αγωγή που ακολουθεί ο χρήστης.
- Ο πίνακας μέτρησης **meal** περιέχει εγγραφές που αφορούν στα γεύματα του χρήστη.
- Ο πίνακας μέτρησης **glycated_hemoglobin** περιέχει εγγραφές για την τιμή γλυκοζυλιωμένης αιμοσφαιρίνης του χρήστη.
- Ο πίνακας **source** προσδιορίζει αν οι εισαγόμενες τιμές στους παραπάνω πίνακες μετρήσεων έγιναν χειροκίνητα ή μέσω κάποιας πλατφόρμας υγείας (π.χ. Google Fit).



Εικόνα 3.2.1 Το ER Διάγραμμα της βάσης δεδομένων forDiabetes.

Η εξαγωγή των δεδομένων από τη βάση έγινε σε αρχεία μορφής CSV.

3.3 Μετασχηματισμοί Δεδομένων

Όλα τα δεδομένα προς ανάλυση είναι καταχωρημένα στους πίνακες που περιγράφονται στην 3.2 ενότητα. Από κατασκευής των πινάκων έχει εξασφαλιστεί ότι δεν υπάρχουν διπλότυπες εγγραφές.

3.3.1 Χαρακτηριστικά των χρηστών της εφαρμογής

Στη βάση δεδομένων στον πίνακα user_profile υπάρχουν 23.305 εγγεγραμμένοι χρήστες από όλο τον κόσμο. Όπως είναι φυσικό καλύπτουν ένα ευρύ φάσμα προσωπικών χαρακτηριστικών όπως τον τύπο σακχαρώδους διαβήτη που πάσχουν, την ηλικία αλλά και τη χώρα τους.

Για να εξεταστεί και να παρουσιαστεί η ποικιλία των χαρακτηριστικών που εμφανίζεται μεταξύ των χρηστών, μετασχηματίστηκαν τα δεδομένα του πίνακα user_profile και δημιουργήθηκε ο Πίνακας 3.3.1.1.

Για τη συνολική και πιο «καθαρή» απεικόνιση της κατανομής των χρηστών ως προς τη χώρα που δήλωσαν δημιουργήθηκε στη βάση δεδομένων ένας επιπλέον πίνακας με το όνομα continent. Σκοπός του πίνακα continent είναι να αντιστοιχεί τη χώρα του πίνακα user_profile στην ήπειρο που ανήκει.

Η τελική μορφή του πίνακα user_profile απεικονίζεται στον Πίνακα 3.3.1.1:

User_Id	Country	Age	Continent	Diabetes_type
6977	MX	57	NAm	40
6978	US	51	NAm	20
6979	FR	36	EU	10
6980	US	60	NAm	20
6981	IT	61	EU	20
6982	US	45	NAm	50
6983	GR	71	EU	20
6984	NULL	NULL	NULL	NULL
6985	BR	41	SA	40
6986	GB	22	EU	10
6987	DE	32	EU	NULL
6988	NULL	NULL	NULL	NULL
6989	EC	47	SA	10
6990	PT	47	EU	10
6991	US	40	NAm	NULL
6993	AU	43	OC	40
6994	DE	61	EU	20
6995	AR	24	SA	20
6996	DE	54	EU	20

Πίνακας 3.3.1.1 Στιγμιότυπο περιεχομένου του πίνακα user_profile.

Τα πεδία του Πίνακα 3.3.1.1 αναλύονται ως εξής:

- **User_Id:** Αριθμητικό αναγνωριστικό, μοναδικό για κάθε χρήστη. Αποτελεί τον αύξοντα αριθμό στη βάση δεδομένων και δεν υπάρχει προφανής σύνδεση με το ονοματεπώνυμο του χρήστη που αντιπροσωπεύει.
- **Country:** Η χώρα που δήλωσε ο χρήστης.
- **Age:** η ηλικία του χρήστη. Υπολογίζεται βάσει του έτους γέννησης.
- **Continent:** Προσδιορίζει σε ποια ήπειρο ανήκει ο χρήστης βάσει της χώρας που δήλωσε.
- **Diabetes_type:** Ο τύπος Σακχαρώδους Διαβήτη του χρήστη.

Όπως εξηγείται στη συνέχεια, το αριθμητικό περιεχόμενο της στήλης Age διακριτοποιείται προς διευκόλυνση της αναλυτικής επεξεργασίας.

3.3.2 Μετρήσεις της εφαρμογής

Η εφαρμογή *forDiabetes* δίνει τη δυνατότητα στον πάσχοντα να καταγράφει σε καθημερινή βάση (και μάλιστα περισσότερες φορές για την ίδια μέρα) μέτρησης δεικτών που επηρεάζουν την πορεία της πάθησης του. Δείκτες όπως την καθημερινή αθλητική άσκηση, τιμές γλυκόζης όπως αυτές μετρώνται σε διαφορετικές στιγμές της ημέρας, τους υδατάνθρακες που καταναλώνονται στο κάθε ένα γεύμα, τη λήψη φάρμακων, την αρτηριακή πίεση και το βάρος του. Όλα αυτά αποθηκεύονται σε διαφορετικούς πίνακες στη βάση των δεδομένων.

Φυσικά κάθε χρήστης καταχωρεί τιμές με διαφορετική συχνότητα στους παραπάνω πίνακες μετρήσεων. Η τελική μορφή κάθε πίνακα μέτρησης για να εντοπιστούν και να αναλυθούν οι συνεπείς χρήστες είναι ο εξής:

User_Id	Registrations	Frequency	Time
2	211	0.24	896
3	254	0.28	893
4	87	0.10	881
6	22	0.44	50
7	330	0.37	893
8	322	1.02	315
9	6	2.00	3
12	31	0.12	268
13	106	0.35	303
14	524	0.98	537
17	1	NA	0
18	3	1.50	2
19	1	NA	0
23	15	0.07	229
25	59	0.16	367
26	5	2.50	2

Πίνακας 3.3.2.1 Στιγμιότυπο ενός πίνακα μέτρησης για την συνολική δράση κάθε χρήστη.

Τα πεδία του Πίνακα 3.3.2.1 αναλύονται ως εξής:

- **User_Id:** Αριθμητικό αναγνωριστικό, μοναδικό για κάθε χρήστη.
- **Registrations:** το σύνολο των καταχωρήσεων που έχει πραγματοποιήσει ο χρήστης στο συγκεκριμένο πίνακα μέτρησης.
- **Time:** το διάστημα που ήταν ενεργός ο χρήστης στον πίνακα μέτρησης . Πιο συγκεκριμένα υπολογίζεται από την χρονική διαφορά μεταξύ της πρώτης καταχώρησης του χρήστη στο πίνακα με την τελευταία. Σε κάθε γραμμή του πίνακα μέτρησης υπάρχει το πεδίο `created_at` όπου αποθηκεύεται η ημερομηνία που έγινε η συγκεκριμένη καταχώρηση. Μονάδα μέτρησης είναι η μέρα.
- **Frequency:** είναι η συχνότητα με την οποία ο χρήστης καταχωρεί δεδομένα στον αντίστοιχο πίνακα μέτρησης. Προκύπτει από τη διαίρεση $\text{Registrations} / \text{Time}$.

Συνοπτικά η πληροφορία που περιγράφει ο Πίνακας 3.3.2.1 είναι η συνολική δράση κάθε χρήστη που καταχώρησε δεδομένα στον πίνακα μέτρησης, ο χρόνος που μεσολάβησε μεταξύ της πρώτης και τελευταίας καταχώρησης και τη συχνότητα με την οποία καταχωρούσε τιμές.

Κεφάλαιο 4ο: ΔΙΕΡΕΥΝΗΤΙΚΗ ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ: ΑΠΟΤΕΛΕΣΜΑΤΑ

Στο κεφάλαιο αυτό γίνεται η παρουσίαση των αποτελεσμάτων που προέκυψαν μέσα από τη διερευνητική ανάλυση των δεδομένων. Περιλαμβάνει τους αλγόριθμους παραγωγή γραφημάτων που επιτρέπουν την απόκτηση εικόνας για την κατανομή των χρηστών βάσει των χαρακτηριστικών τους. Ακολουθεί η παρουσίαση του αλγορίθμου για τον εντοπισμό του στατιστικού δείγματος των συνεπών χρηστών για κάθε τύπο μέτρησης ξεχωριστά αλλά και σε ζευγάρια, ο οποίος εστιάζει τόσο στο σύνολο των ημερών που ήταν ενεργός ένας χρήστης στη μέτρηση όσο και τη συχνότητα με την οποία καταχωρεί δεδομένα. Έπειτα, ακολουθεί η ανάλυση και οπτικοποίηση των χαρακτηριστικών των στατιστικών δειγμάτων που προέκυψαν.

4.1 Κατανομή των χρηστών βάσει των χαρακτηριστικών τους

Για την ανάδειξη της κατανομής των χαρακτηριστικών των χρηστών, του Πίνακα 3.3.1.1 `user_profile`, παράγονται διαφορετικές γραφικές απεικόνιση για κάθε γνώρισμα.

4.1.1 Κατανομή αριθμού χρηστών ανά τύπο Σακχαρώδους Διαβήτη

Για την παρουσίαση της κατανομής των χρηστών ως προς το τύπο Σακχαρώδους Διαβήτη (ΣΔ) που πάσχουν δημιουργήθηκε ένα ραβδόγραμμα. Στον οποίο στον άξονα x αναπαριστώνται οι διαφορετικές τιμές που έχει λάβει η μεταβλητή που δηλώνει τον τύπο ΣΔ (`Diabetes_type`) και στον άξονα y αναπαριστάται το σύνολο των χρηστών που ανήκουν στο συγκεκριμένο τύπο ΣΔ.

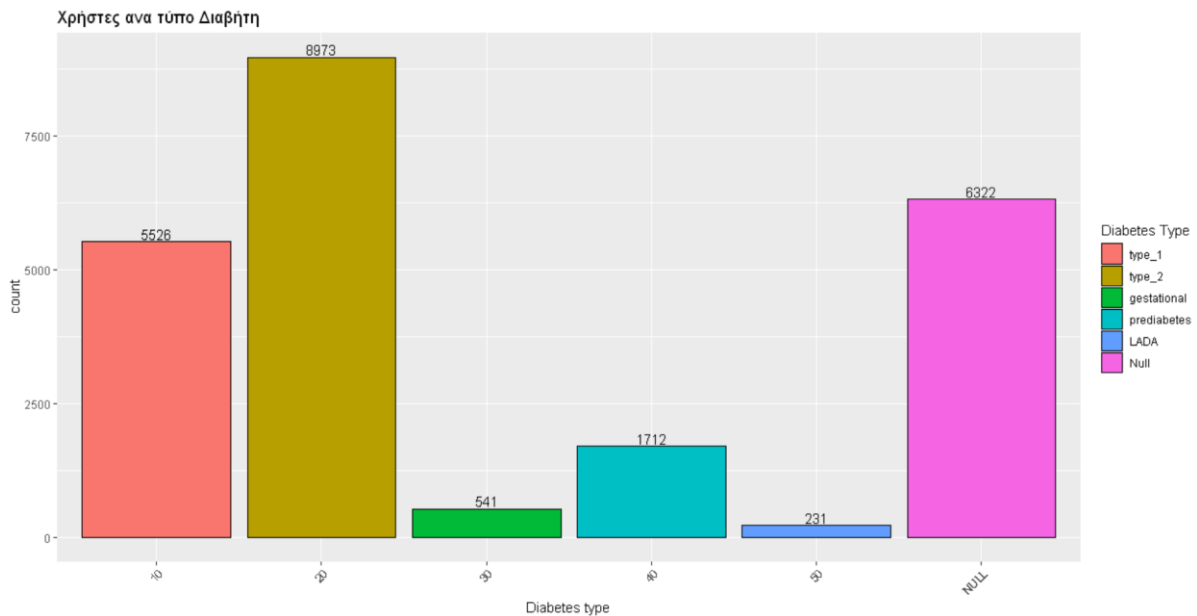
Ο αλγόριθμος που παράγει τη γραφική απεικόνιση έχει ως εξής:

- Για τη μεταβλητή `Diabetes_type` του πίνακα `user_profile`
 - Υπολογίζεται το πλήθος των χρηστών που είχαν `Diabetes_type=10` που αντιστοιχεί στο ΣΔ Τύπου 1 (`type_1`).
 - Υπολογίζεται το πλήθος των χρηστών που είχαν `Diabetes_type=20` που αντιστοιχεί στο ΣΔ Τύπου 2(`type_2`).
 - Υπολογίζεται το πλήθος των χρηστών που είχαν `Diabetes_type=30` που αντιστοιχεί στο ΣΔ κύησης (`gestational`).
 - Υπολογίζεται το πλήθος των χρηστών που είχαν `Diabetes_type=40` που αντιστοιχεί στο προδιαβήτη (`prediabetes`).
 - Υπολογίζεται το πλήθος των χρηστών που είχαν `Diabetes_type=50` που αντιστοιχεί στο Λανθάνων Αυτοάνοσο ΣΔ (`LADA`).
 - Υπολογίζεται το πλήθος των χρηστών που είχαν `Diabetes_type=NULL` που αντιστοιχεί στους χρήστες που δε δήλωσαν τον τύπο διαβήτη που πάσχουν.

Στην Εικόνα 4.1.1.1 παρουσιάζεται το γράφημα, που αντιστοιχεί στο σύνολο των χρηστών για κάθε τύπο Σακχαρώδους Διαβήτη (ΣΔ). Παρατηρείται ότι:

- Οι περισσότεροι χρήστες φέρουν να πάσχουν από ΣΔ τύπου 2, συνολικά 8.973 και αποτελούν το 38,5% των χρηστών της εφαρμογής.
- Στο ΣΔ τύπου 1 ανήκουν 5.526 χρήστες, συνιστούν το 23,71% των χρηστών.

- Στον ΣΔ κήσης ανήκουν 541 χρήστες, συνιστούν το 2,23% των χρηστών.
- Στον προδιαβήτη ανήκουν 1712 χρήστες, συνιστούν το 7,34% των χρηστών.
- Στον ΣΔ LADA ανήκουν 231, συνιστούν το 0,99% των χρηστών
- Μεγάλος αριθμός χρηστών δεν έχει δηλώσει το ΣΔ που πάσχουν, συγκεκριμένα 6322 και συνιστούν το 27,12% των χρηστών.



Εικόνα 4.1.1.1 Κατανομή αριθμού χρηστών ανά τύπο διαβήτη.

4.1.2 Κατανομή αριθμού χρηστών ανά χώρα

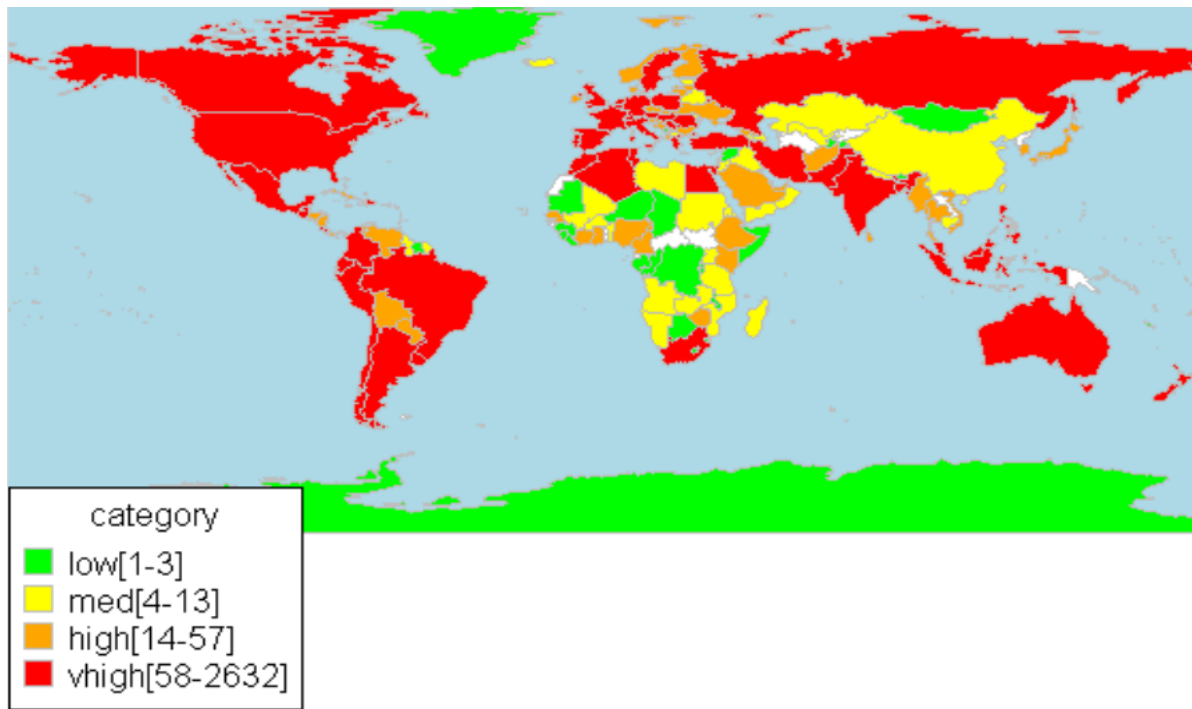
Για την παρουσίαση της κατανομής των χρηστών ως προς τη χώρα που δήλωσαν δημιουργήθηκε ένα `worldmap` γράφημα. Το συγκεκριμένο γράφημα εξυπηρετεί στη καλύτερη απεικόνιση δεδομένων παγκόσμιας κλίμακας.

Ο αλγόριθμος που παράγει το συγκεκριμένο γράφημα έχει ως εξής:

- Για κάθε χώρα στο πίνακα `user_profile`
 - Υπολογίζεται το σύνολο των χρηστών που τη δήλωσε
 - Καταχωρείται σε ένα `data frame` το ζεύγος «χώρα» - «σύνολο χρηστών».
- Για κάθε ζευγάρι τιμών «χώρα» - «σύνολο χρηστών» του `data frame`
 - Εισάγονται οι τιμές στο χάρτη
 - Καθορίζεται το όνομα της στήλης που περιέχει το αναγνωριστικό της χώρας.
 - Καθορίζεται ο τύπος κώδικα που χρησιμοποιείται για το όνομα της χώρας, "ISO3" για κωδικούς γραμμάτων ISO 3 ή "UN" για αριθμητικό κωδικό χώρας.
 - Καταχωρούνται όλα τα στοιχεία σε ένα νέο `data frame` τύπου `SpatialPolygonsDataFrame` (sPDF).
- Για κάθε αντικείμενο του sPDF `data frame`

- Εμφανίζονται στο χάρτη
 - Προσδιορίζεται το όνομα της στήλης που περιέχει τα δεδομένα που θα σχεδιαστούν στο χάρτη, στην προκειμένη περίπτωση το «σύνολο χρηστών».

Στην Εικόνα 4.1.2.1 παρουσιάζεται το γράφημα που αντιστοιχεί στο σύνολο των χωρών που έχουν δηλώσει οι χρήστες. Οι χώρες χρωματίζονται στο χάρτη βάσει της ελάχιστης τιμής των χρηστών σε μια χώρα ($\min = 1$ χρήστης = 0%) και της μέγιστης τιμής ($\max = 2.632$ χρήστες=100%). Χωρίζονται σε τέσσερις ισοπληθής ομάδες/κατηγορίες. Στην κατηγορία low που αναπαριστάται με πράσινο χρώμα στο γράφημα, ανήκουν οι χώρες που το σύνολο των χρηστών που την έχει δηλώσει είναι κάτω του 25%. Αντίστοιχα με κίτρινο χρώμα αναπαριστάται η κατηγορία med, στην οποία ανήκουν οι χώρες που το σύνολο των χρηστών της είναι πάνω από το 25% έως το 50%. Στην κατηγορία high ανήκουν οι χώρες στις οποίες το πλήθος των χρηστών της είναι πάνω από το 50% έως το 75%. Τέλος είναι η κατηγορία vhigh, στην οποία ανήκουν οι χώρες που το πλήθος των χρηστών της είναι πάνω από 75% έως το 100%. Οι χώρες με λευκό χρώμα δεν αντιστοιχούν σε κανένα χρήστη. Παρατηρώντας την Εικόνα 4.1.2.1 άξιο αναφοράς αποτελεί το γεγονός πως μια ελληνική εφαρμογή χρησιμοποιείται από την παγκόσμια αγορά υγείας.



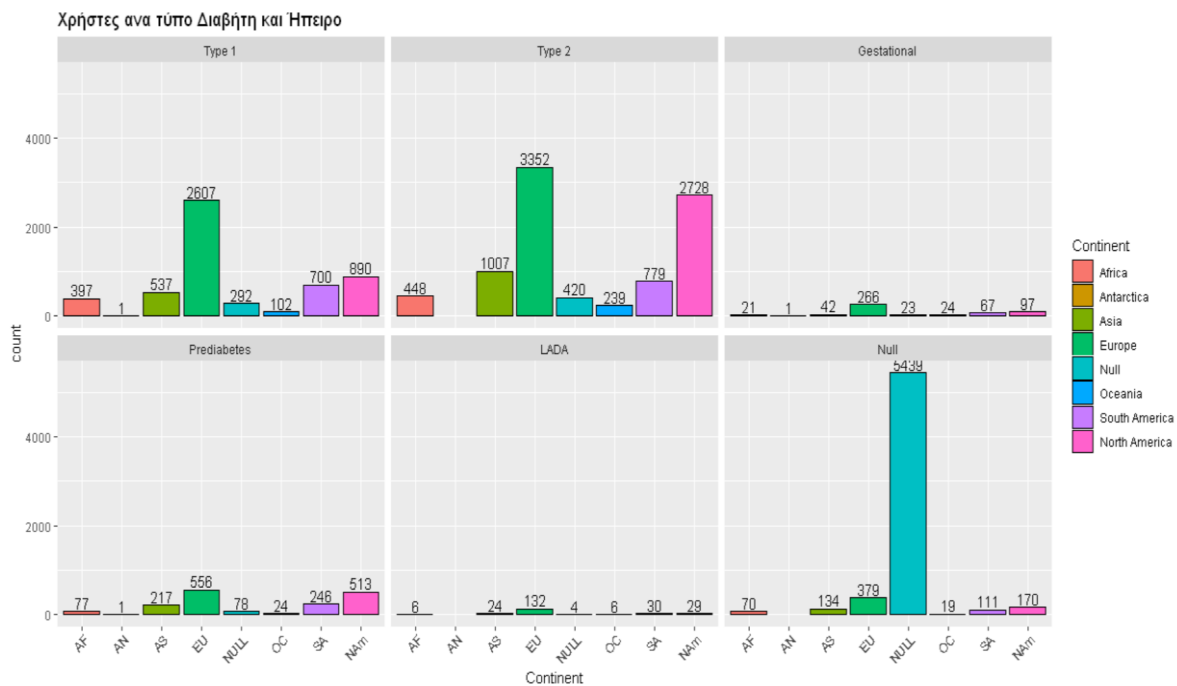
Εικόνα 4.1.2.1 Κατανομή αριθμού χρηστών με βάση τη χώρα που δήλωσαν στο πίνακα user_profile.

4.1.3 Κατανομή αριθμού χρηστών ανά ήπειρο

Από την Εικόνα 4.1.2.1 μπορεί κανείς να καταλάβει ότι ο αριθμός των χωρών που συμμετέχουν στην εφαρμογή είναι αρκετά μεγάλος. Έτσι για κάθε τύπο Σακχαρώδους Διαβήτη για την απεικόνιση της κατανομής του αριθμού των χρηστών ως προς τη χώρα που δήλωσαν έγινε με βάση σε ποια ήπειρο ανήκει. Δημιουργήθηκε για κάθε τύπο Σακχαρώδους Διαβήτη ένα ραβδόγραμμα, το οποίο στον άξονα x αναπαριστώνται οι διαφορετικές τιμές που έχει λάβει η μεταβλητή που προσδιορίζει την ήπειρο (Continent) και στον άξονα y αναπαριστάται το σύνολο των χρηστών που ανήκουν στη συγκεκριμένη ήπειρο.

Ο αλγόριθμος που παράγει το γράφημα έχει ως εξής:

- Για κάθε τύπο Σακχαρώδους Διαβήτη
 - Για τη μεταβλητή Continent
 - Υπολογίζεται το σύνολο των χρηστών που έχουν Continent = AF που αντιστοιχεί στην Αφρική.
 - Υπολογίζεται το σύνολο των χρηστών που έχουν Continent = AN που αντιστοιχεί στην Ανταρκτική.
 - Υπολογίζεται το σύνολο των χρηστών που έχουν Continent = AS που αντιστοιχεί στην Ασία
 - Υπολογίζεται το σύνολο των χρηστών που έχουν Continent = EU που αντιστοιχεί στην Ευρώπη.
 - Υπολογίζεται το σύνολο των χρηστών που έχουν Continent = OC που αντιστοιχεί στην Ωκεανία.
 - Υπολογίζεται το σύνολο των χρηστών που έχουν Continent = SA που αντιστοιχεί στη Νότια Αμερική.
 - Υπολογίζεται το σύνολο των χρηστών που έχουν Continent = NAm που αντιστοιχεί στη Βόρεια Αμερική.
 - Υπολογίζεται το σύνολο των χρηστών που έχουν Continent = NULL που αντιστοιχεί στους χρήστες που δε δήλωσαν χώρα.



Εικόνα 4.1.3.1 Κατανομή αριθμού χρηστών ανά ήπειρο και τύπο διαβήτη.

Στην Εικόνα 4.1.3.1 απεικονίζεται το γράφημα, που αντιστοιχεί στην ήπειρο στην οποία ανήκουν οι χρήστες με βάση τη χώρα που δήλωσαν. Παρατηρείται ότι :

- Η κατάταξη των ηπειρών ως προς το σύνολο των χρηστών έχει ως εξής:

- Το μεγαλύτερο ποσοστό των χρηστών σε όλους τους τύπου σακχαρώδους διαβήτη είναι από την Ευρώπη, αποτελούν το 31,28% του συνόλου.
 - Ακολουθεί η Βόρεια Αμερική με ποσοστό 18,99%
 - Η Ασία με ποσοστό 8,41%
 - Η Νότια Αμερική με ποσοστό 8,29%
 - Αφρική με ποσοστό 4,37%
 - Ωκεανία με ποσοστό 1,77%
 - Ανταρκτική με ποσοστό 0,012%
- Υπάρχει μεγάλος αριθμός χρηστών που δεν έχουν δηλώσει τη χώρα στα στοιχεία του λογαριασμού τους,, αποτελούν το 26,84% των χρηστών.

4.1.4 Κατανομή αριθμού χρηστών ανά ηλικία

Για την παρουσίαση της κατανομής των χρηστών ως προς την ηλικία, στο πίνακα user_profile δημιουργήθηκε μια επιπλέον μεταβλητή AgeLevel. Συγκεκριμένα έγινε διακριτικοποίηση της μεταβλητής Age σε κατηγορηματικές τιμές που δηλώνουν την ηλικιακή ομάδα στην οποία ανήκει ο χρήστης.

Πιο αναλυτικά οι διακριτές τιμές της μεταβλητής AgeLevel είναι :

- Infant: για Age στο διάστημα [0,1]
- Child: για Age στο διάστημα [2,18]
- Adult: για Age στο διάστημα [19,59]
- Senior: για Age στο διάστημα [60,89]
- Nonagenarian: για Age στο διάστημα [90,99]

Για τη γραφική απεικόνιση της κατανομής χρηστών ανά ηλικιακή ομάδα δημιουργήθηκε για κάθε τύπο Σακχαρώδους Διαβήτη ένα ραβδόγραμμα. Σε αυτό στον άξονα x αναπαριστώνται οι διαφορετικές τιμές που έχει λάβει η μεταβλητή AgeLevel και στον άξονα y αναπαριστάται το σύνολο των χρηστών που ανήκουν στη συγκεκριμένη ηλικιακή ομάδα.

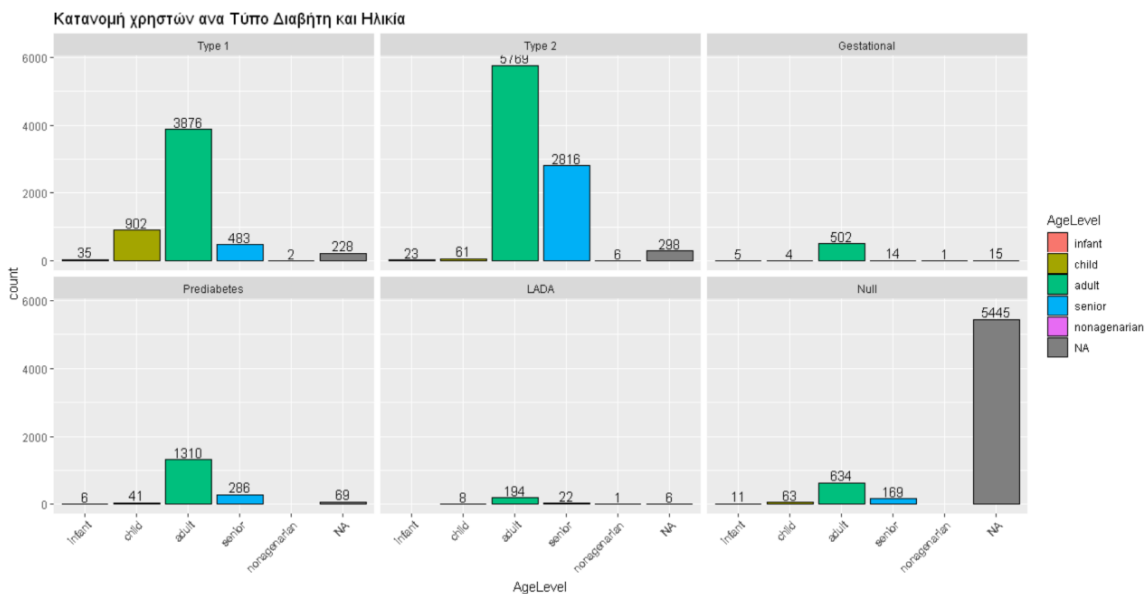
Ο αλγόριθμος που παράγει το γράφημα έχει ως εξής:

- Για κάθε τύπο Σακχαρώδους Διαβήτη
 - Για τη μεταβλητή AgeLevel
 - Υπολογίζεται το σύνολο των χρηστών που έχουν AgeLevel = infant που αντιστοιχεί στα βρέφη
 - Υπολογίζεται το σύνολο των χρηστών που έχουν AgeLevel = child που αντιστοιχεί στα παιδιά.
 - Υπολογίζεται το σύνολο των χρηστών που έχουν AgeLevel = adult που αντιστοιχεί στους ενήλικες
 - Υπολογίζεται το σύνολο των χρηστών που έχουν AgeLevel = senior που αντιστοιχεί στους ενήλικες

- Υπολογίζεται το σύνολο των χρηστών που έχουν AgeLevel = nonagenarian που αντιστοιχεί στους χρήστες άνω των 90 ετών.
- Υπολογίζεται το σύνολο των χρηστών που έχουν AgeLevel = NA που αντιστοιχεί στους χρήστες που δεν δήλωσαν έτος γέννησης.

Στην Εικόνα 4.1.4.1 παρουσιάζεται το γράφημα, που αντιστοιχεί στην ηλικιακή ομάδα στην οποία ανήκουν οι χρήστες με βάση το έτος γέννησης που δήλωσαν. Παρατηρείται ότι:

- Η κατάταξη των ηλικιακών ομάδων βάσει του συνόλου των χρηστών που τις απαρτίζουν έχει ως εξής:
 - Οι περισσότεροι χρήστες είναι ενήλικες, το 52,71%
 - ακολουθούν οι υπερήλικες με ποσοστό 16,26%,
 - τα παιδιά με ποσοστό 4,62%,
 - τα βρέφη με ποσοστό 0,34%,
 - οι άνω των 90 ετών με ποσοστό 0,042%
- Υπάρχει μεγάλο μέρος των χρηστών που δεν έχει δηλώσει το έτος γέννησης του, αντιστοιχεί στο 26%
- Στην περίπτωση του διαβήτη κύησης (gestational) πέρα από τους ενήλικες οι υπόλοιποι 39 χρήστες είναι αδύνατο να ανήκουν στο συγκεκριμένο τύπο ΣΔ.



Εικόνα 4.1.4.1 Κατανομή αριθμού χρηστών ανά ηλικιακή ομάδα και τύπο διαβήτη.

4.2 Αξιόπιστες περιπτώσεις χρηστών στις μετρήσεις της εφαρμογής

Για κάθε τύπο μέτρησης της εφαρμογής υπολογίζεται το σύνολο των συνεπών χρηστών και αναλύονται ως προς τα χαρακτηριστικά τους. Συνεπείς χαρακτηρίζονται οι χρήστες που ανταπεξέρχονται συστηματικά στις υποχρεώσεις τους, στην συγκεκριμένη περίπτωση να μετρούν και καταχωρούν τιμές στην εφαρμογή. Τα δεδομένα τους μπορούν να αναλυθούν και να προκύψει κάποιο μοντέλο/μοτίβο εξάρτησης μεταξύ των διαφορετικών τύπων μετρήσεων. Η συνέπεια των χρηστών προέκυψε βάσει δυο γενικών κριτηρίων.

4.2.1 Κριτήρια Αξιολόγησης

Για να αποκλειστούν οι περιπτώσεις όπου ο χρήστης είτε έχει μικρό διάστημα στο οποίο είναι ενεργός στον πίνακα μέτρησης (της μορφής Πίνακα 3.3.2.1) και κατ' επέκταση έχει λίγες καταχωρήσεις, είτε είναι μεγάλο χρονικό διάστημα ενεργός και έχει λίγες καταχωρήσεις ορίστηκαν τα ακόλουθα κριτήρια.

Το πρώτο κριτήριο αξιολόγησης αφορά το σύνολο των ημερών που ήταν ενεργός ο χρήστης στον κάθε τύπο μέτρησης πίνακα (Πίνακας 3.3.2.1). Πραγματοποιείται φιλτράρισμα στο πίνακα μέτρησης και απορρίπτονται οι χρήστες που είχαν χρόνο (πεδίο Time) μικρότερο των πενήντα (50) ημερών. Οι πενήντα (50) ημέρες και άνω θεωρούνται εύλογο χρονικό διάστημα για να παρατηρηθεί η συνέπεια του χρήστη στη πάροδο του χρόνου.

Το δεύτερο κριτήριο αφορά την συχνότητα με την οποία ο χρήστης καταχωρεί δεδομένα στον κάθε τύπο μέτρησης πίνακα (Πίνακας 3.3.2.1). Συνεπείς χαρακτηρίστηκαν οι χρήστες που είχαν συχνότητα καταχωρήσεων (πεδίο Frequency) μεγαλύτερο ή ίσο από 0,7 ανά ημέρα. Αυτό σημαίνει ότι ο χρήστης έχει συνοχή στις καταχωρήσεις του ενώ παράλληλα του δίνεται το περιθώριο της αδράνειας για κάποιες ημέρες.

4.2.2 Αλγόριθμος για την παραγωγή γραφημάτων

Δέχεται διαφορετικό σύνολο δεδομένων για κάθε τύπο μέτρησης. Ο πίνακας έχει την μορφή όπως τον Πίνακα 3.3.1.1.

Ο πρώτος αλγόριθμος αφορά την απεικόνιση των αποτελεσμάτων των κριτηρίων αξιολόγησης και έχει ως εξής:

- Για κάθε τύπο μέτρησης
 - Εύρεση των χρηστών με συνολικό χρόνο δράσης στη μέτρηση στο διάστημα [0,50).
 - Υπολογίζεται το σύνολο των χρηστών
 - Δημιουργείται πίνακας με τα πεδία «χρήστες» - «σύνολο χρηστών»
 - Εύρεση των χρηστών με συνολικό χρόνο δράσης στη μέτρηση στο διάστημα [50,50+].
 - Υπολογίζεται το σύνολο των χρηστών
 - Δημιουργείται πίνακας με τα πεδία «χρήστες» - «σύνολο χρηστών»
 - Καταχωρούνται σε ένα data frame οι χρήστες
 - Ενώνονται οι δυο πίνακες που περιέχουν τα ζευγάρια τιμών και δημιουργείται ένα data frame «κατηγορία» - «σύνολο χρηστών».
- Για κάθε ζευγάρι του data frame «κατηγορία» - «σύνολο χρηστών».
 - Υπολογίζεται το επί τις % ποσοστό για την κάθε κατηγορία, ως προς το σύνολο των χρηστών
 - Υπολογίζεται το μερικό άθροισμα κάθε κατηγορίας

- Στο data frame περιέχει του χρήστες με Time στο διάστημα [50,50+]
 - Εύρεση των χρηστών με συχνότητα καταχωρήσεων στη μέτρηση στο διάστημα [0,0.7)
 - Υπολογίζεται το σύνολο των χρηστών
 - Δημιουργείται πίνακας με τα πεδία «χρήστες» - «σύνολο χρηστών»
 - Εύρεση των χρηστών με συχνότητα καταχωρήσεων στη μέτρηση στο διάστημα [0.7,1)
 - Καταχωρούνται οι χρήστες σε ένα data frame
 - Υπολογίζεται το σύνολο των χρηστών
 - Δημιουργείται πίνακας με τα πεδία «χρήστες» - «σύνολο χρηστών»
 - Ενώνονται οι δυο πίνακες που περιέχουν τα ζευγάρια τιμών και δημιουργείται ένα data frame «κατηγορία» - «σύνολο χρηστών»
- Για κάθε ζευγάρι του data frame «κατηγορία» - «σύνολο χρηστών»
 - Υπολογίζεται το επί τις % ποσοστό της κάθε κατηγορίας , ως προς στο σύνολο των χρηστών
 - Υπολογίζεται το μερικό άθροισμα κάθε κατηγορίας

Ο επόμενος αλγόριθμος αφορά την απεικόνιση των χαρακτηριστικών των χρηστών που αποτελούν το στατιστικό δείγμα των συνεπών χρηστών κάθε μέτρησης.

Το data frame που περιέχει τους χρήστες με συχνότητα εγγραφής/καταχωρήσεων στο διάστημα [0.7, 1], έγινε ένωση με το data frame που περιέχει τα χαρακτηριστικά των χρηστών. Πιο αναλυτικά, έχοντας σαν κύριο κλειδί το αναγνωριστικό του χρήστη, έγινε αντιστοίχιση στο data frame που περιέχει τα χαρακτηριστικά τους. Το αποτέλεσμα της ένωσης, δημιουργήθηκε ένα data frame με τα πεδία «αναγνωριστικό χρήστη» - «τύπος ΣΔ» - «ηλικιακή ομάδα». Ο αλγόριθμος που υλοποιήθηκε για την απεικόνιση την κατανομής των χαρακτηριστικών των συνεπών χρηστών για κάθε τύπο μέτρησης είναι:

- Στο data frame που περιέχει την τριπλέτα «αναγνωριστικό χρήστη» - «τύπος ΣΔ» - «ηλικιακή ομάδα»
 - Για κάθε τύπο Σακχαρώδους Διαβήτη
 - Υπολογίζεται το σύνολο χρηστών για κάθε ηλικιακή ομάδα.

4.2.3 Απεικόνιση Γραφημάτων

Για τη γραφική απεικόνιση της κατανομής των χρηστών ως προς το κριτήριο του συνολικού χρόνου που ήταν ενεργός στο συγκεκριμένο τύπο μέτρησης, δημιουργήθηκε ένα donut διάγραμμα. Ουσιαστικά πρόκειται για ένα ραβδόγραμμα το οποίο γίνεται κυκλικό χάρη στο πολικό σύστημα συντεταγμένων (coord_polar()). Το αντίστοιχο διάγραμμα δημιουργήθηκε για την παρουσίαση της κατανομής των χρηστών ως προς το κριτήριο της συχνότητας με την οποία καταχωρούν τιμές στο συγκεκριμένο τύπο μέτρησης.

Για τη γραφική απεικόνιση της κατανομής των συνεπών χρηστών ως προς τα χαρακτηριστικά τους (τύπος ΣΔ και ηλικία) δημιουργήθηκε ένα ραβδόγραμμα που στο άξονα x αναπαριστάται ο τύπος ΣΔ

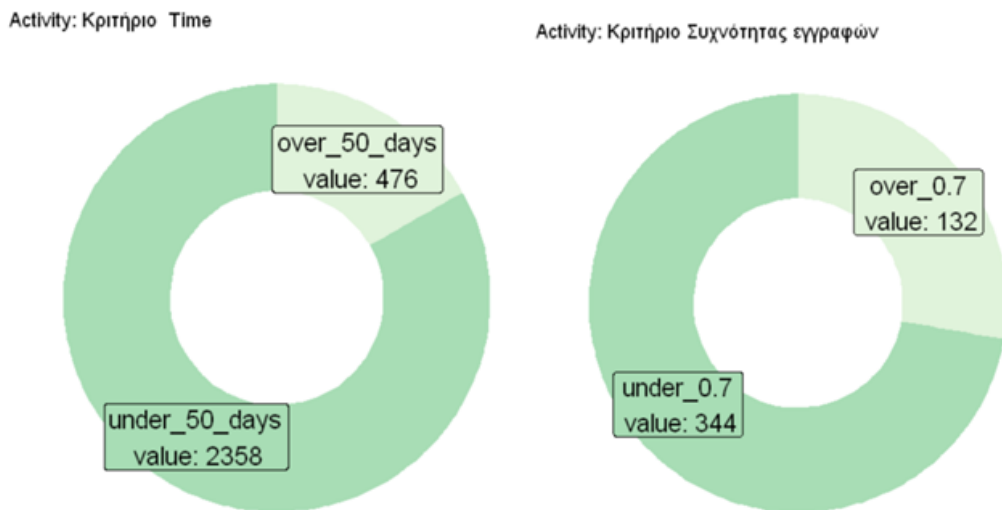
και στον άξονα y το σύνολο τους. Η απεικόνιση της κατανομής της ηλικιακής ομάδας των χρηστών σε κάθε τύπο ΣΔ γίνεται με τη βοήθεια των χρωμάτων.

4.2.4 Μέτρηση Activity

Στο πίνακα μέτρησης activity στη βάση δεδομένων, αποθηκεύονται όλα τα δεδομένα που περιγράφουν τη σωματική άσκηση του χρήστη. Περιλαμβάνει πεδία που αφορούν το είδος της άσκησης που έκανε (περπάτημα, τρέξιμο, κολύμβηση είναι μερικά παραδείγματα ανάμεσα σε πολλά), τη διάρκεια της άσκησης, πόσες θερμίδες έκαψε, την απόσταση που διένυσε σε μέτρα, τη διάρκεια της άσκησης σε λεπτά αλλά και την ημερομηνία της καταχώρησης.

Στον πίνακα activity στη βάση των δεδομένων υπάρχουν συνολικά 103.328 εγγραφές και πραγματοποιήθηκαν από 2.834 χρήστες. Εφαρμόζοντας το κριτήριο του χρόνου (πεδίο Time \geq 50 ημερών) και της συχνότητας εγγραφών ανά ημέρα (πεδίο Frequency \geq 0,7) παράγονται τα γραφήματα της Εικόνας 4.2.4.1.

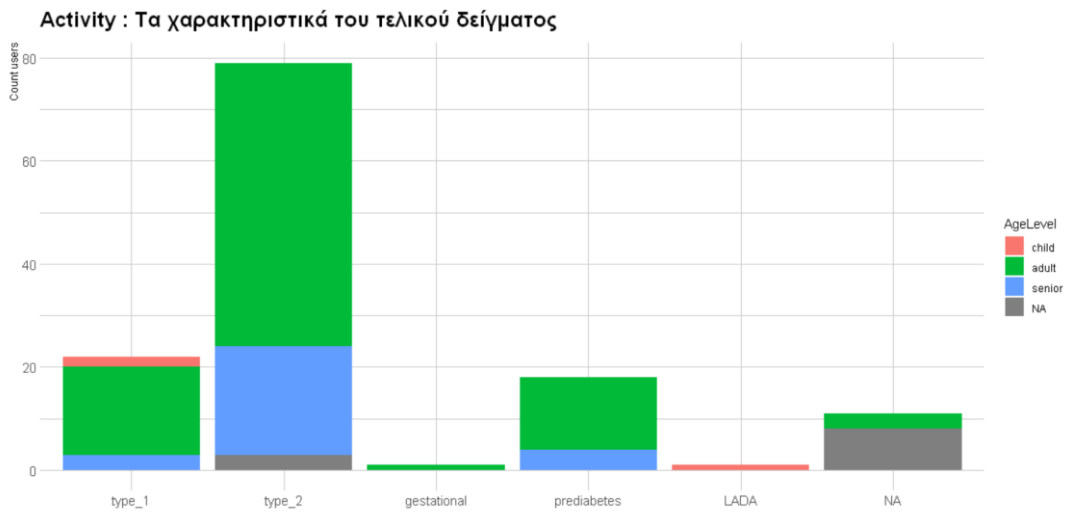
Στη Εικόνα 4.2.4.1 απεικονίζονται τα γραφήματα, που παρουσιάζουν τα αποτελέσματα από την εφαρμογή των δυο φίλτρων συνέπειας στη μέτρηση Activity. Παρατηρείται από το γράφημα (α) ότι μόνο 476 (16,8%) χρήστες έχουν καταχωρήσει δεδομένα με συνολική χρονική περίοδο πάνω από πενήντα (50) ημέρες ενώ 2.358 (83,2%) χρήστες κάτω από πενήντα (50) ημέρες. Από τους 476 χρήστες, από το (β) φαίνεται ότι οι 132 (27,8%) χρήστες είχαν συχνότητα εγγραφών/καταχωρήσεων τιμών μεγαλύτερη από 0,7 ανά ημέρα ενώ οι υπόλοιποι 344 (72,2%) είχαν μικρότερη συχνότητα καταχωρήσεων. Έτσι το τελικό στατιστικό δείγμα είναι 132 χρήστες και συνιστούν το 4,66% των χρηστών που έχουν καταχωρήσει δεδομένα στη μέτρηση activity.



Εικόνα 4.2.4.1 Φίλτρα συνέπειας χρήστη στη μέτρηση Activity: (α) χρονική περίοδος \geq 50, (β) συχνότητα καταχωρήσεων \geq 0.7.

Από το γράφημα της Εικόνας 4.2.4.2, που παρουσιάζει τα χαρακτηριστικά του στατιστικού δείγματός των συνεπών χρηστών της μέτρησης activity φαίνεται ότι οι χρήστες διαβήτη ΣΔ τύπου 2 καταχωρούν πιο συστηματικά δεδομένα της σωματικής άσκησης και μάλιστα οι ενήλικες και υπερήλικες. Ακολουθούν οι ενήλικες ΣΔ τύπου 1 και οι προδιαβητικοί. Τόσο ο ΣΔ κήσης όσο και ο αυτοάνοσος

ΣΔ αποτελούνται από ένα άτομο. Υπάρχουν βέβαια και οι χρήστες που δεν είναι γνωστό ούτε ο τύπος ΣΔ τους, ούτε η ηλικιακή τους ομάδα.

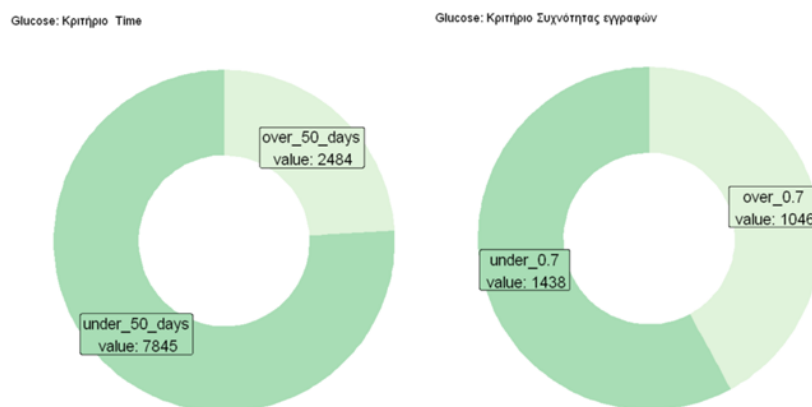


Εικόνα 4.2.4.2 Τα χαρακτηριστικά των συνεπών χρηστών στη μέτρηση Activity.

4.2.5 Μέτρηση Glucose

Στη μέτρηση glucose αποθηκεύονται όλα τα δεδομένα που καταχωρούν οι χρήστες σχετικά με το επίπεδο της γλυκόζης στο αίμα τους. Η τιμή της μέτρηση μπορεί να έγινε μετά από νηστεία, πριν ή μετά από γεύμα, πριν τον ύπνο ή κάτι άλλο. Υπάρχει πεδίο που προσδιορίζει αν η μέτρηση έγινε πριν ή μετά από σωματική άσκηση. Τέλος αποθηκεύεται η ημερομηνία που έγινε η καταχώρηση.

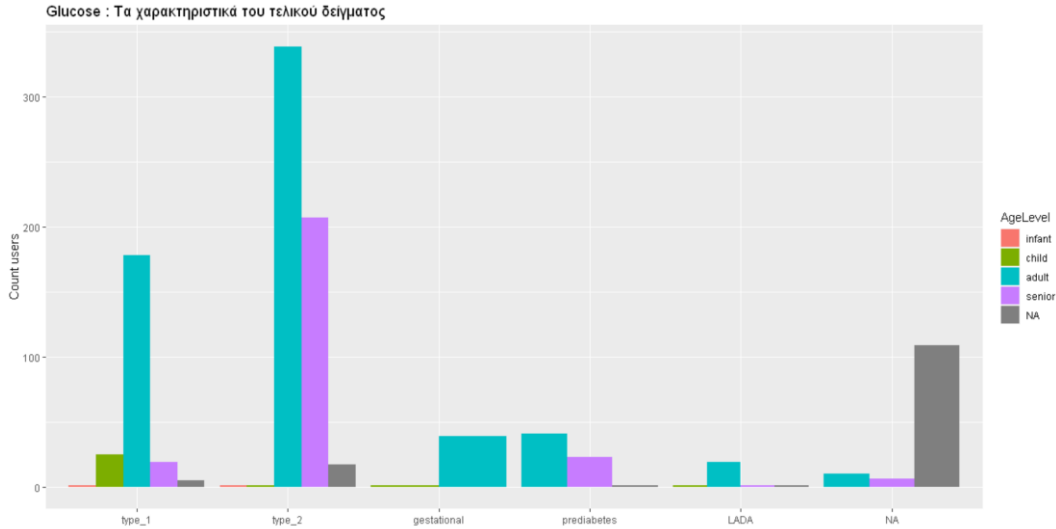
Στο πίνακα glucose στην βάση δεδομένων υπάρχουν 737.158 εγγραφές και πραγματοποιήθηκαν από 10.329 χρήστες. Εφαρμόζοντας τα φίλτρα συνέπειας στη μέτρηση Glucose, τα αποτελέσματα παρουσιάζονται στην Εικόνα 4.2.5.1.



Εικόνα 4.2.5.1 Φίλτρα συνέπειας χρήστη στη μέτρηση Glucose: (α) χρονική περίοδος ≥ 50 , (β) συχνότητα καταχωρήσεων ≥ 0.7

Στη Εικόνα 4.2.5.1 απεικονίζονται τα γραφήματα, που παρουσιάζουν τα αποτελέσματα από την εφαρμογή των δυο φίλτρων συνέπειας στη μέτρηση glucose. Παρατηρείται από το (α) ότι μόνο 2.484

(24%) χρήστες έχουν καταχωρήσει δεδομένα με συνολική χρονική περίοδο πάνω από πενήντα (50) ημέρες, Από τους 2484 χρήστες, από το (β) φαίνεται ότι μόνο οι 1046 (42,1%) καταχωρούν τιμές στη μέτρηση glucose με συχνότητα ίση ή μεγαλύτερη από 0,7 ανά ημέρα. Το τελικό στατιστικό δείγμα είναι 1.046 χρήστες και συνιστούν το 10,1% των χρηστών που έχουν καταχωρήσει δεδομένα στη μέτρηση glucose.

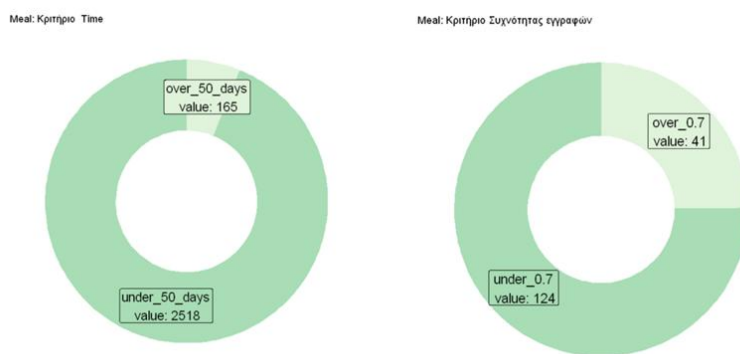


Εικόνα 4.2.5.2 Τα χαρακτηριστικά των συνεπών χρηστών στη μέτρηση Glucose.

Από το γράφημα της Εικόνας 4.2.5.2 που παρουσιάζει τα χαρακτηριστικά των συνεπών χρηστών στη μέτρηση glucose φαίνεται ότι οι πλειοψηφία των χρηστών που καταχωρούν πιο συχνά δεδομένα στο συγκεκριμένο τύπο μέτρησης είναι οι διαβητικοί ΣΔ τύπου 2. Η ηλικιακή ομάδα στην οποία ανήκουν είναι η “ενήλικες” και ακολουθούν η “υπερήλικες”. Στο στατιστικό δείγμα των συνεπών χρηστών υπάρχουν και αρκετοί ενήλικες ΣΔ τύπου 1 και χρήστες που δεν είναι γνωστό κανένα στοιχείο τους. Όλες οι υπόλοιπες περιπτώσεις στις ομάδες ΣΔ έχουν κάτω από 50 χρήστες.

4.2.6 Μέτρηση Meal

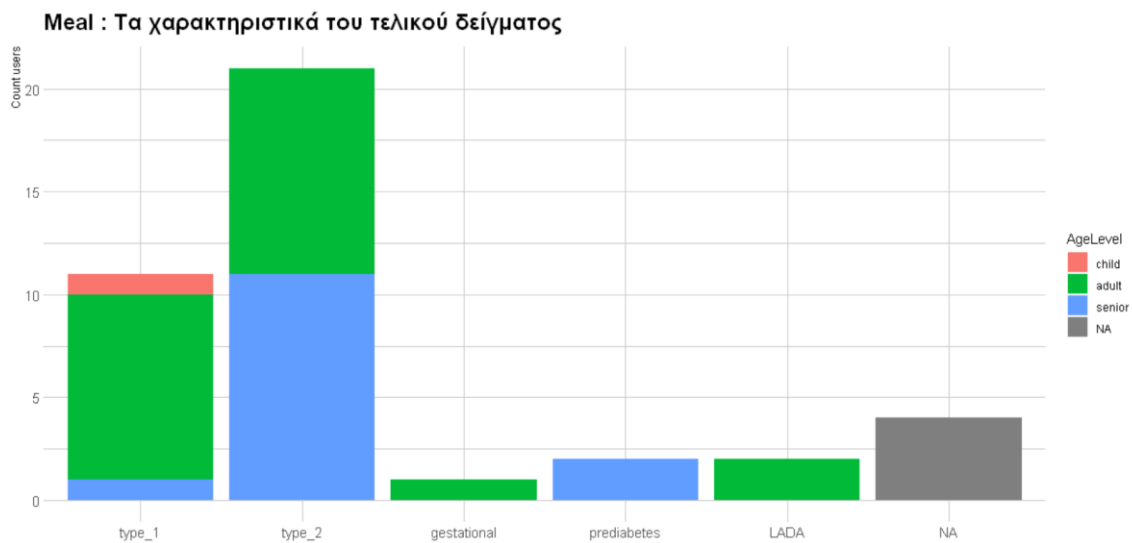
Τα δεδομένα του χρήστη σχετικά με το γεύμα που κατανάλωσε αποθηκεύονται στο πίνακα meal στη βάση δεδομένων. Περιλαμβάνει πεδία που αφορούν τους υδατάνθρακες που κατανάλωσε κατά την διάρκεια του γεύματος και αν ο γλυκαιμικός δείκτης του ήταν χαμηλός, μέτριος ή υψηλός.



Εικόνα 4.2.6.1 Φίλτρα συνέπειας χρήστη στη μέτρηση Meal: (α) χρονική περίοδος ≥ 50 , (β) συχνότητα καταχωρήσεων ≥ 0.7

Στον πίνακα meal στη βάση δεδομένων υπάρχουν συνολικά 26.640 εγγραφές από 2.683 χρήστες. Εφαρμόζοντας τα φίλτρα συνέπειας στη μέτρηση Meal, παράγονται τα γραφήματα της Εικόνας 4.2.6.1 στα οποία αποτυπώνονται τα αποτελέσματα.

Στη Εικόνα 4.2.6.1 απεικονίζονται τα γραφήματα, που παρουσιάζουν τα αποτελέσματα από την εφαρμογή των δυο φίλτρων συνέπειας στη μέτρηση meal. Άξιο λόγου αποτελεί το γεγονός ότι από το γράφημα (α) μόνο 165 (6,1%) χρήστες καταχωρούν δεδομένα πάνω από πενήντα (50) ημέρες στη μέτρηση meal. Από τους 165 χρήστες, από το (β) παρατηρείται ότι μόνο 41 (25%) χρήστες καταχωρούν δεδομένα στη μέτρηση meal με συχνότητα ίση ή μεγαλύτερη από 0,7 ανά ημέρα. Το τελικό στατιστικό δείγμα των συνεπών χρηστών της μέτρησης meal αποτελείται από 41 χρήστες και συνιστούν το 1,52% των χρηστών που έχουν καταχωρήσει δεδομένα



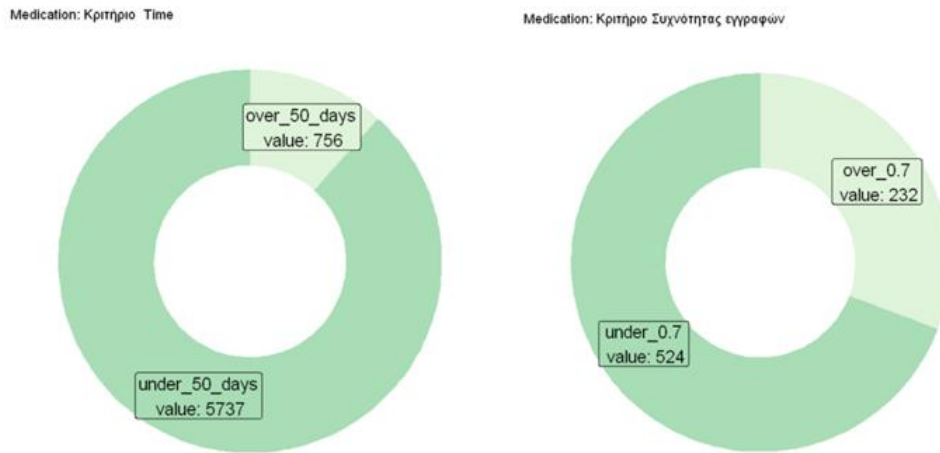
Εικόνα 4.2.6.2 Τα χαρακτηριστικά των συνεπών χρηστών στη μέτρηση Meal.

Από το γράφημα της Εικόνας 4.2.6.2 που παρουσιάζουν τα χαρακτηριστικά των συνεπών χρηστών στη μέτρηση meal φαίνεται ότι οι πλειοψηφία των χρηστών που καταχωρούν πιο συχνά δεδομένα στο συγκεκριμένο τύπο μέτρησης είναι οι διαβητικοί ΣΔ τύπου 2 (51,22%) και είναι μόνο ενήλικες (47,62%) και υπερήλικες (52,38%). Υπάρχουν και 11 χρήστες διαβητικοί τύπου ΣΔ 1 (26,83%), στους οποίους είναι και το μοναδικό παιδί του στατιστικού δείγματος. Οι χρήστες στους υπόλοιπους ΣΔ τύπους είναι ελάχιστοι, συνολικά αποτελούνται από 5 (12,19%) χρήστες. Επιπλέον, υπάρχουν και 4 χρήστες που δεν γνωρίζουμε ούτε τον τύπο ΣΔ αλλά ούτε την ηλικιακή ομάδα.

4.2.7 Μέτρηση Medication

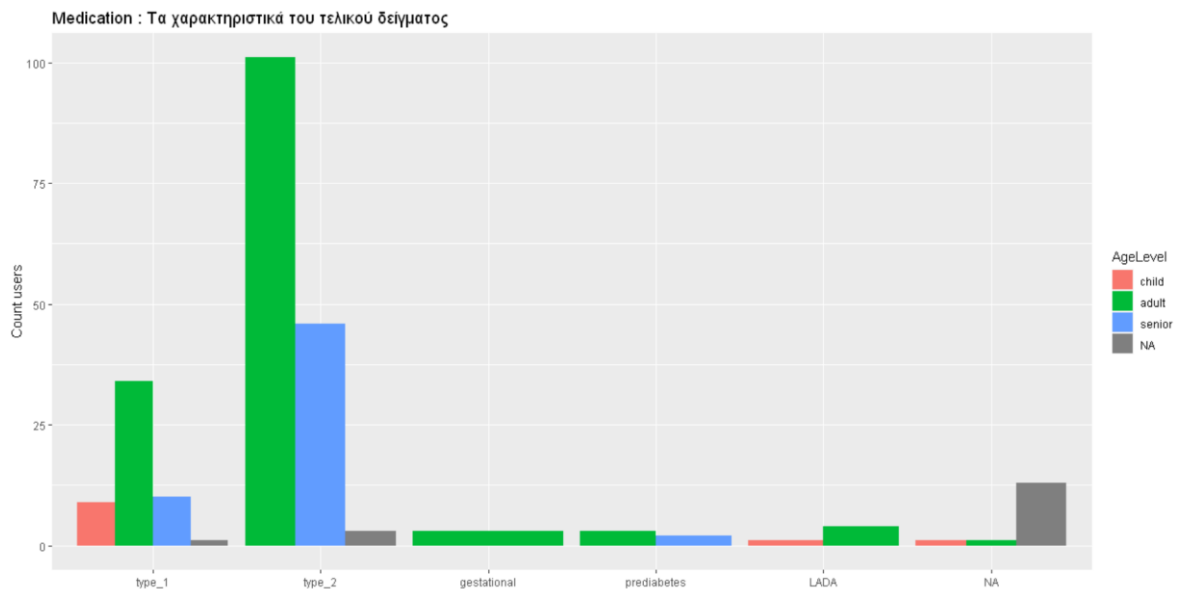
Στη μέτρηση medication αποθηκεύονται τα δεδομένα των χρηστών που αφορούν την φαρμακευτική αγωγή που ακολουθούν. Πιο συγκεκριμένα περιέχει πεδία για το όνομα και την ποσότητα του φαρμάκου που χορηγήθηκε. Ακόμα, προσδιορίζει αν το φάρμακο ήταν κάποιο χάπι, ινσουλίνη ή κάποιο άλλο ενέσιμο φάρμακο. Φυσικά υπάρχει πεδίο με την ημερομηνία της καταχώρησης.

Στον πίνακα medication στη βάση δεδομένων υπάρχουν 162.467 εγγραφές και πραγματοποιήθηκαν από 6.493 χρήστες. Εφαρμόζοντας τα φίλτρα συνέπειας στη μέτρηση Medication, παράγονται τα γραφήματα της Εικόνας 4.2.7.1 στα οποία αποτυπώνονται τα αποτελέσματα.



Εικόνα 4.2.7.1 Φίλτρα συνέπειας χρήστη στη μέτρηση Medication: (α) χρονική περίοδος ≥ 50 , (β) συχνότητα καταχωρήσεων ≥ 0.7 .

Στην Εικόνα 4.2.7.1 απεικονίζονται τα γραφήματα, που παρουσιάζουν τα αποτελέσματα από την εφαρμογή των δυο φίλτρων συνέπειας στη μέτρηση medication. Παρατηρείται από το γράφημα (α) μόνο 756 (12,6%) χρήστες καταχωρούν δεδομένα πάνω από πενήντα (50) ημέρες στη μέτρηση medication. Από τους 756 χρήστες, από το (β) παρατηρείται ότι μόνο 232 (30,6%) χρήστες καταχωρούν δεδομένα στη μέτρηση medication με συχνότητα ίση ή μεγαλύτερη από 0,7 ανά ημέρα. Το στατιστικό δείγμα των συνεπών χρηστών της μέτρησης medication αποτελείται από 232 χρήστες και συνιστούν ποσοστό 3,57% των χρηστών που έχουν καταχωρήσει δεδομένα στο συγκεκριμένο τύπο μέτρησης.



Εικόνα 4.2.7.2 Τα χαρακτηριστικά των συνεπών χρηστών στη μέτρηση Medication.

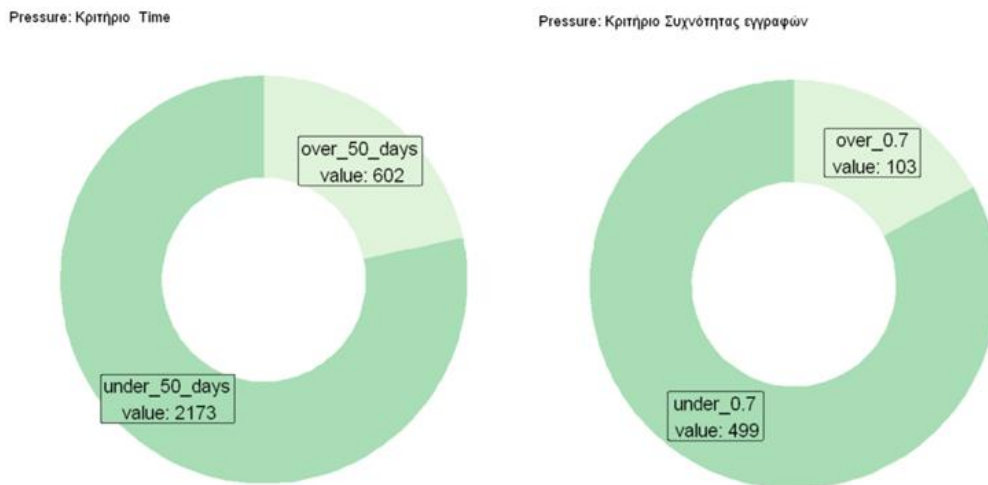
Από το γράφημα της Εικόνας 4.2.7.2, που παρουσιάζουν τα χαρακτηριστικά των συνεπών χρηστών στη μέτρηση medication, προκύπτει γρήγορα το συμπέρασμα ότι υπερисχύουν αριθμητικά οι χρήστες ΣΔ τύπου 2 (64,65%) και ακολουθούν αριθμητικά οι χρήστες ΣΔ τύπου 1 (23,27%). Στον ΣΔ τύπο 2 υπάρχουν 101 ενήλικες (67,33%), 46 υπερήλικες (30,67%) και 3 (2%) χρήστες άγνωστης ηλικιακής ομάδας. Στον ΣΔ τύπο 1 υπάρχουν 9 παιδιά (16,67%), 34 ενήλικες (62,96%), 10 υπερήλικες (18,52)

και 1 χρήστης άγνωστης ηλικιακής ομάδας. Οι υπόλοιποι τύποι ΣΔ, συμπεριλαμβανομένου και του άγνωστου τύπου ΣΔ, αποτελούνται συνολικά από 28 χρήστες (12,08) . Από τους 28 χρήστες ,για τους 13 δεν είναι γνωστό κανένα χαρακτηριστικό τους.

4.2.8 Μέτρηση Pressure

Η μέτρηση pressure περιέχει όλα τα δεδομένα που έχουν σχέση με την αρτηριακή πίεση των χρηστών. Πιο συγκεκριμένα διαθέτει πεδία για τη διαστολική και συστολική πίεση. Ακόμη, υπάρχει πεδίο για το σφυγμό (χτύπους καρδιάς ανά λεπτό) και την ημερομηνία της καταχώρησης.

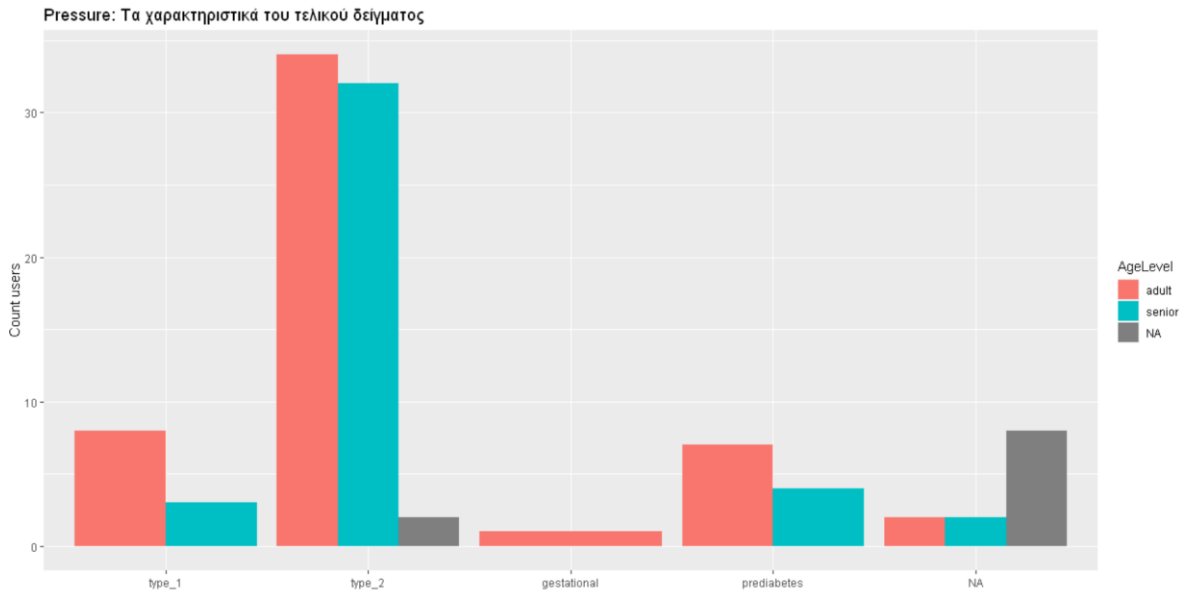
Στη μέτρηση pressure στη βάση δεδομένων υπάρχουν συνολικά 57.600 εγγραφές και πραγματοποιήθηκαν από 2.775 χρήστες.



Εικόνα 4.2.8.1 Φίλτρα συνέπειας χρήστη στη μέτρηση Pressure: (α) χρονική περίοδος ≥ 50 , (β) συχνότητα καταχωρήσεων ≥ 0.7 .

Στη Εικόνα 4.2.8.1 απεικονίζονται τα γραφήματα, που παρουσιάζουν τα αποτελέσματα από την εφαρμογή των δυο φίλτρων συνέπειας στη μέτρηση pressure. Παρατηρείται από το γράφημα (α) μόνο 602 (21,69%) χρήστες καταχωρούν δεδομένα πάνω από πενήντα (50) ημέρες στη μέτρηση pressure και υπόλοιποι 2.173 (78,31) χρήστες έχουν μικρότερο διάστημα ημερών. Από τους 602 χρήστες, από το (β) παρατηρείται ότι μόνο 103 (17,1%) χρήστες καταχωρούν δεδομένα στη μέτρηση medication με συχνότητα ίση ή μεγαλύτερη από 0,7 ανά ημέρα ενώ οι υπόλοιποι 499(82,9%) με μικρότερη συχνότητα. Το στατιστικό δείγμα των συνεπών χρηστών της μέτρησης pressure αποτελείται από 103 χρήστες και συνιστούν το 3,71% των χρηστών που έχουν καταχωρήσει δεδομένα στο συγκεκριμένο τύπο μέτρησης.

Από το γράφημα της Εικόνας 4.2.8.2, που παρουσιάζει τα χαρακτηριστικά των συνεπών χρηστών στη μέτρηση pressure, προκύπτει γρήγορα το συμπέρασμα ότι υπερисχύουν αριθμητικά οι χρήστες ΣΔ τύπου 2 (66,02%). Στον ΣΔ τύπο 2 υπάρχουν 34 ενήλικες (50%), 32 υπερήλικες (47,06%) και 2 (2,94%) χρήστες άγνωστης ηλικιακής ομάδας. Στον ΣΔ τύπο 1, όπως και στο προδιαβήτη, υπάρχουν 11 χρήστες και είναι ενήλικες και υπερήλικες. Στον ΣΔ κήσης και στον άγνωστο τύπο ΣΔ οι χρήστες συνολικά είναι 18 και από αυτούς για τους 8 δεν είναι γνωστό κανένα χαρακτηριστικό τους.



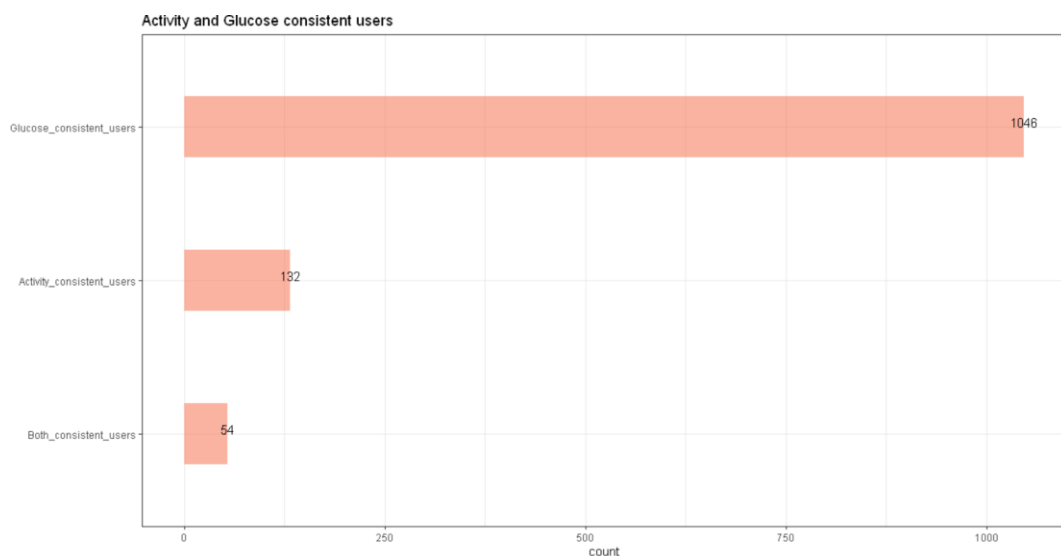
Εικόνα 4.2.8.2 Τα χαρακτηριστικά των συνεπών χρηστών στη μέτρηση Pressure.

4.3 Ζευγάρια Μετρήσεων

Μετά τον εντοπισμό των χρηστών που είναι συνεπείς σε κάθε τύπο μέτρησης επόμενο βήμα είναι να βρεθούν και σε ζευγάρια μετρήσεων. Συγκεκριμένα υπολογίζεται για κάθε τύπο μέτρησης της βάσης δεδομένων οι κοινοί συνεπείς χρήστες με τη βασική μέτρηση glucose. Το κάθε στατιστικό δείγμα που προκύπτει θα μπορεί να χρησιμοποιηθεί για να ελεγχθεί η σχέση εξάρτησης διαφορετικών μετρήσεων με τη γλυκόζη.

4.3.1 Μετρήσεις Activity – Glucose

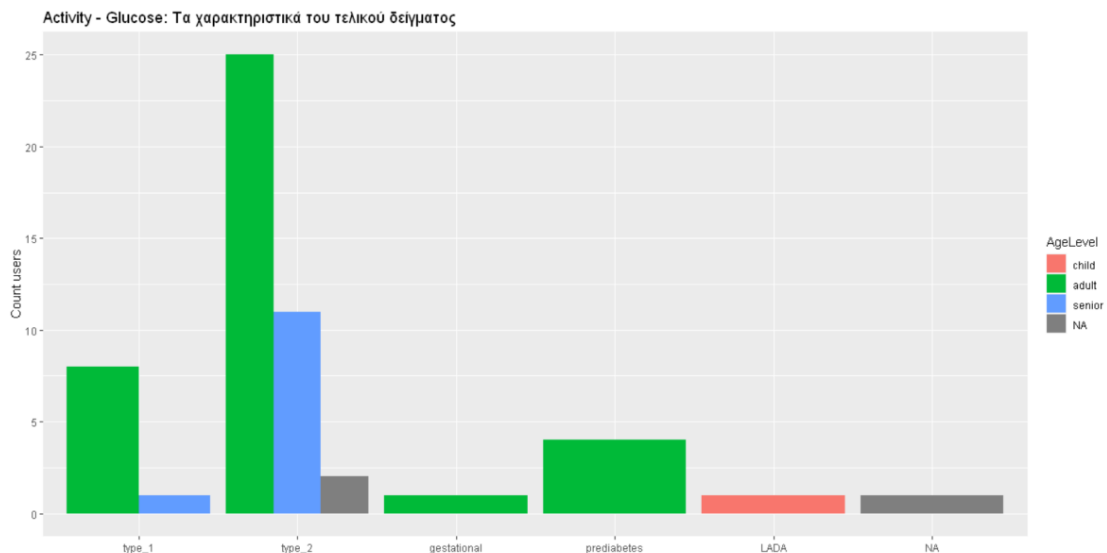
Η συστηματική σωματική άσκηση αποτελεί έναν από τους πιο αποτελεσματικούς αντιμετώπισης του σακχαρώδους διαβήτη. Κατά τη διάρκεια της άσκησης, οι μύες καταναλώνουν γλυκόζη για να παράγουν ενέργεια, μειώνοντας τα επίπεδα της γλυκόζης στο αίμα.



Εικόνα 4.3.1.1 Οι συνεπείς χρήστες στις μετρήσεις Activity - Glucose.

Για τον εντοπισμό του στατιστικού δείγματος που είναι συνεπείς τόσο στη μέτρηση activity όσο και στη μέτρηση glucose, πραγματοποιείται συγχώνευση των στατιστικών δειγμάτων που προέκυψαν από την ενότητα 4.2.4 και 4.2.5. Το αποτέλεσμα απεικονίζεται στο γράφημα της Εικόνας 4.3.1.1.

Στο γράφημα της Εικόνας 4.3.1.1, απεικονίζεται το σύνολο των συνεπών χρηστών για κάθε τύπο μέτρησης ξεχωριστά αλλά και τους κοινούς στις μετρήσεις activity και glucose, προκαλεί εντύπωση το γεγονός ότι οι κοινοί συνεπείς χρήστες συνιστούν ποσοστό 5,1% των συνεπών χρηστών της μέτρησης glucose ενώ στη μέτρηση activity το 40,9%. Το στατιστικό δείγμα των κοινών συνεπών χρηστών αποτελείται από 54 χρήστες.



Εικόνα 4.3.1.2 Χαρακτηριστικά των κοινών συνεπών χρηστών στις μετρήσεις Activity – Glucose.

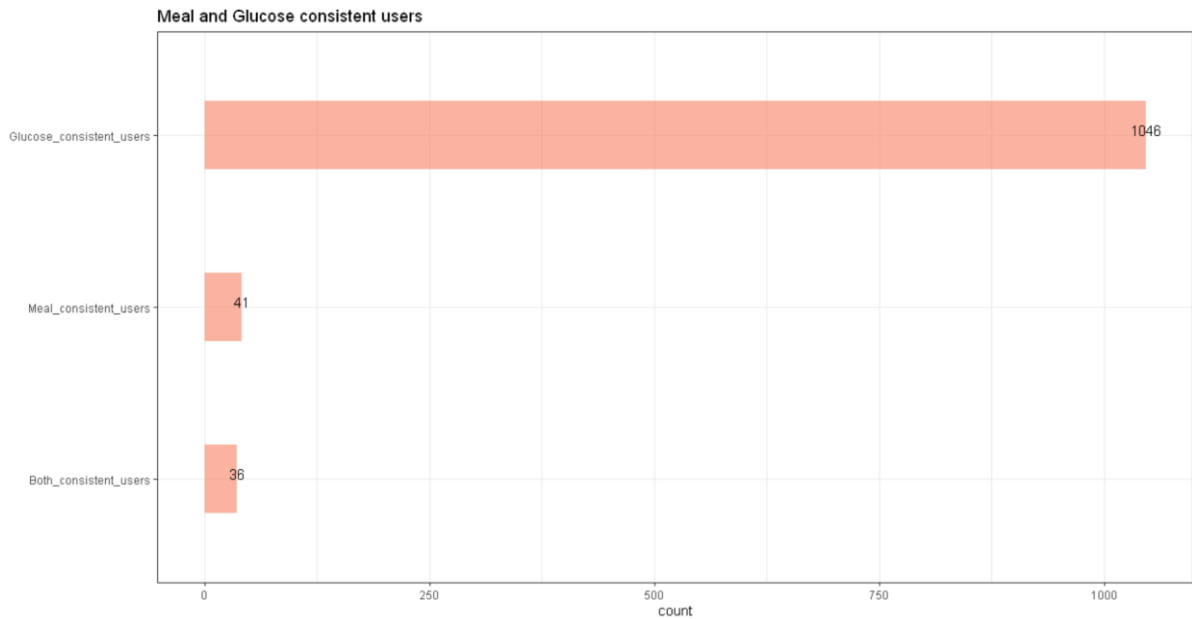
Από το γράφημα της εικόνας 4.3.1.2, που παρουσιάζει τα χαρακτηριστικά του δείγματος που είναι συνεπείς στη μέτρηση activity και glucose, παρατηρείται ότι οι περισσότεροι χρήστες είναι διαβητικοί τύπου 2 (70,37%) και ακολουθούν αριθμητικά οι διαβητικοί τύπου 1 (16,67%). Τόσο ο ΣΔ τύπου 2 όσο και ο ΣΔ τύπου 1 αποτελούνται από ενήλικες και υπερήλικες. Οι υπόλοιποι ΣΔ τύποι και ηλικιακές ομάδες αποτελούνται από ελάχιστους χρήστες. Μάλιστα στο διαβήτη τύπου LADA εκτός ότι αποτελείται από ένα χρήστη είναι και το μοναδικό παιδί στο στατιστικό δείγμα.

4.3.2 Μετρήσεις Meal – Glucose

Ο ρόλος της σωστής διατροφής, τόσο από άποψη ποιότητας όσο και ποσότητας, για τα άτομα που πάσχουν από διαβήτη είναι καθοριστικός και συμβάλλει στη διατήρηση των επιπέδων της γλυκόζης στο αίμα σε φυσιολογικά επίπεδα.

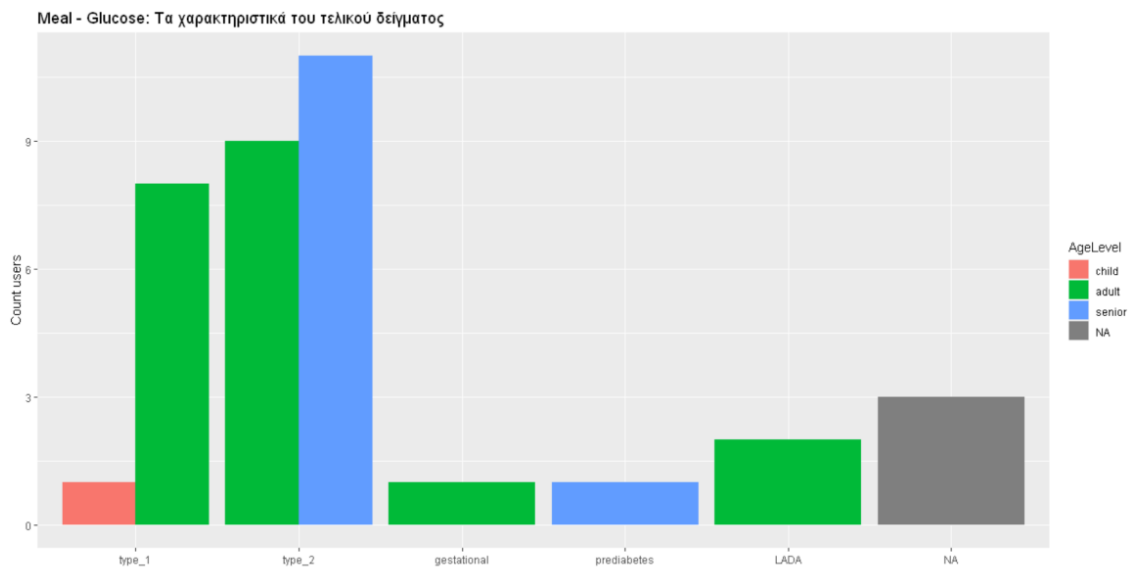
Για τον εντοπισμό του στατιστικού δείγματος που είναι συνεπείς τόσο στη μέτρηση meal όσο και στη μέτρηση glucose, πραγματοποιείται συγχώνευση των στατιστικών δειγμάτων που προέκυψαν από την ενότητα 4.2.6 και 4.2.5. Το αποτέλεσμα απεικονίζεται στο γράφημα της εικόνας 4.3.2.1.

Στο γράφημα της Εικόνας 4.3.2.1, απεικονίζεται το σύνολο των συνεπών χρηστών για κάθε τύπο μέτρησης ξεχωριστά αλλά και τους κοινούς στις μετρήσεις meal και glucose. Ο μικρός αριθμός των κοινών συνεπών χρηστών ήταν αναμενόμενος εξαιτίας του μικρού αριθμού συνεπών χρηστών της μέτρησης meal. Συγκεκριμένα το 87,8% του δείγματος της μέτρησης meal είναι συνεπής και στην μέτρηση glucose. Το γράφημα της εικόνας 4.3.2.2 απεικονίζει τα χαρακτηριστικά του δείγματος που είναι συνεπείς και στους δυο τύπους μετρήσεων.



Εικόνα 4.3.2.1 Οι συνεπείς χρήστες στις μετρήσεις Meal – Glucose.

Από το γράφημα της Εικόνας 4.3.2.2 είναι ευδιάκριτο ότι οι κοινοί χρήστες αποτελούνται από ενήλικες (55,5%) και υπερήλικες (33,3%) κατά το πλείστον. Υπάρχει και ένα παιδί στον ΣΔ τύπου 1 και 3 χρήστες άγνωστου ΣΔ τύπου και ηλικιακής ομάδας. Όσον αναφορά το ΣΔ τύπο, υπερτερεί αριθμητικά ο ΣΔ τύπου 2 με 20 χρήστες (55,5%) και ακολουθεί ο ΣΔ τύπου 1 με 9 χρήστες (25%). Στον προδιαβήτη υπάρχει ένας υπερήλικας χρήστης, στον ΣΔ κύησης μια ενήλικας χρήστης και στον LADA 2 ενήλικες χρήστες. Υπάρχουν 3 χρήστες για τους οποίους δεν είναι γνωστό ούτε ο ΣΔ τύπος ούτε η ηλικιακή ομάδα.



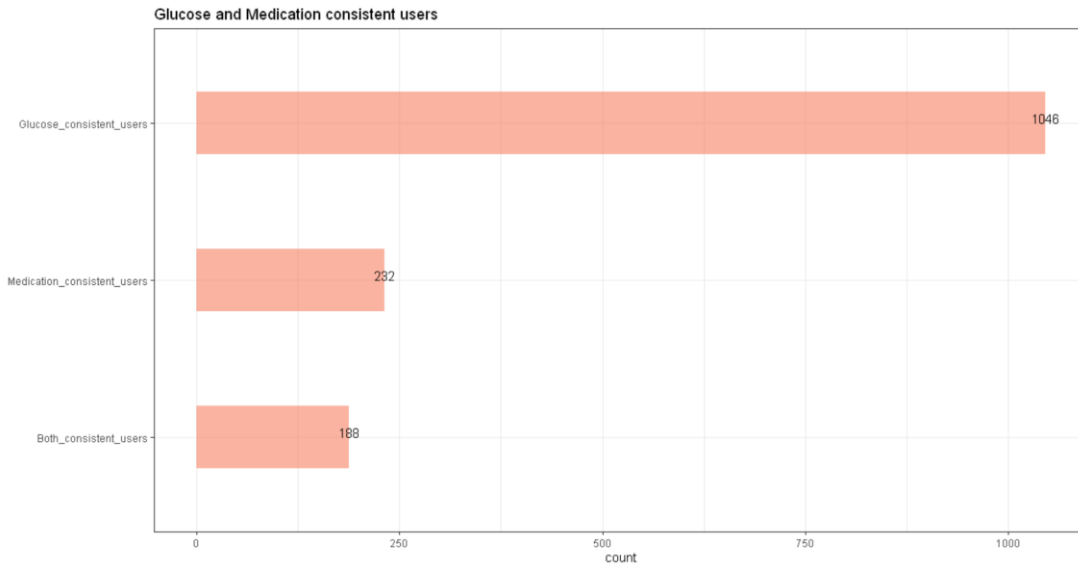
Εικόνα 4.3.2.2 Χαρακτηριστικά των κοινών συνεπών χρηστών στις μετρήσεις Meal – Glucose.

4.3.3 Μετρήσεις Medication – Glucose

Τα άτομα που πάσχουν από οποιαδήποτε μορφή διαβήτη προκειμένου να ελέγξουν τις διαταραχές στα επίπεδα της γλυκόζης στο αίμα τους μπορεί να χρειαστούν κάποια μορφή φαρμακευτικής αγωγής.

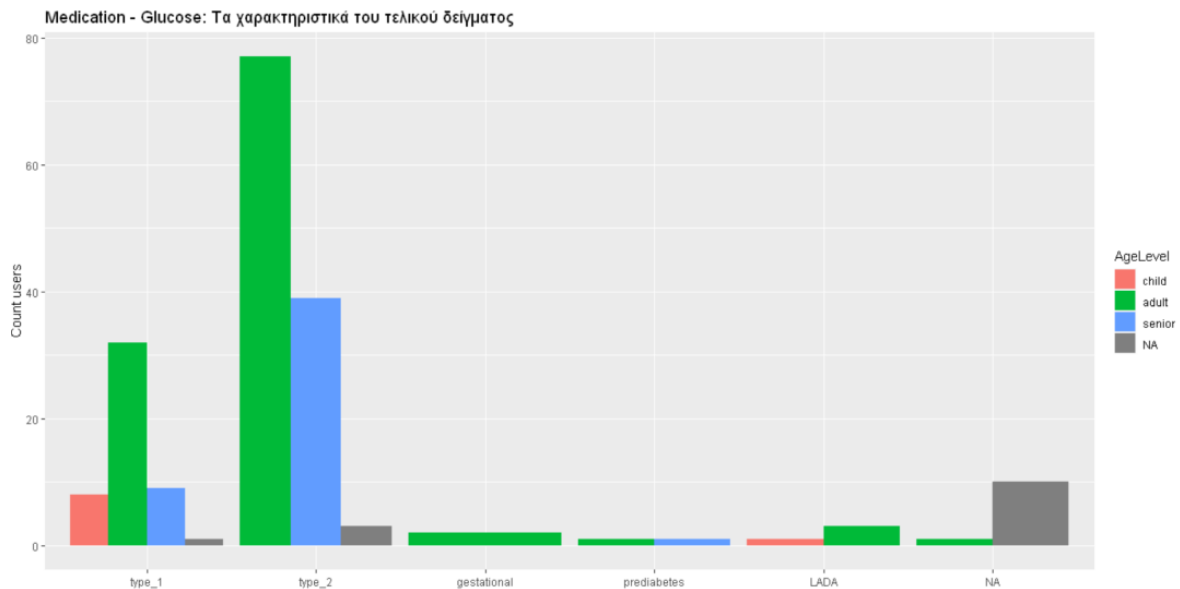
Κεφάλαιο 4

Για τον εντοπισμό του στατιστικού δείγματος που είναι συνεπείς τόσο στη μέτρηση medication όσο και στη μέτρηση glucose, πραγματοποιείται συγχώνευση των στατιστικών δειγμάτων που προέκυψαν από την ενότητα 4.2.7 και 4.2.5. Το αποτέλεσμα απεικονίζεται στο γράφημα της Εικόνας 4.3.3.1.



Εικόνα 4.3.3.1 Οι συνεπείς χρήστες στις μετρήσεις Medication – Glucose.

Σύμφωνα με το γράφημα της Εικόνας 4.3.3.1 οι κοινói συνεπείς χρήστες στις μετρήσεις medication και glucose είναι 188. Συγκεκριμένα το 81,03% του δείγματος της μέτρησης medication είναι συνεπής και στην μέτρηση glucose. Το στατιστικό δείγμα των κοινών συνεπών χρηστών παρουσιάζουν τα χαρακτηριστικά που απεικονίζονται στο γράφημα της Εικόνας 4.3.3.2.



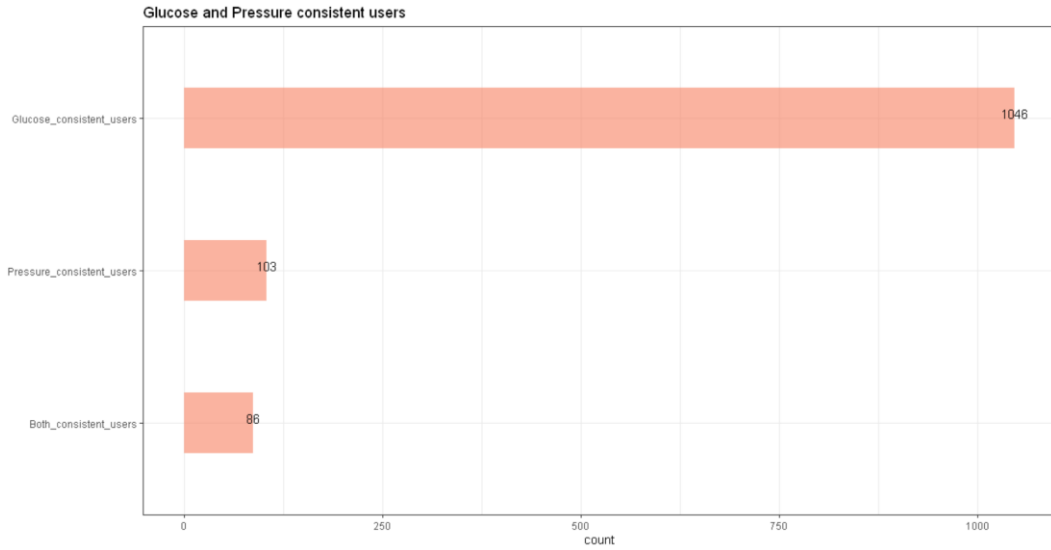
Εικόνα 4.3.3.2 Τα χαρακτηριστικά των κοινών συνεπών χρηστών στις μετρήσεις Medication και Glucose.

Από την Εικόνα 4.3.3.2, οι πλειοψηφία του δείγματος αποτελείται από χρήστες ΣΔ τύπου 2 (63,3%) και ακολουθούν αριθμητικά ο ΣΔ τύπου 1 (26,6%). Πιο αναλυτικά, στον ΣΔ τύπου 2 υπάρχουν 77 ενήλικες (64,7%), 39 υπερήλικες (32,8%) και 3 χρήστες άγνωστης ηλικιακής ομάδας. Στο ΣΔ τύπου 1 υπάρχουν 8 παιδιά (16%), 32 ενήλικες (64%), 9 υπερήλικες (18%) και ένας χρήστης άγνωστης ηλικιακής ομάδας. Οι υπόλοιποι ΣΔ τύποι, συμπεριλαμβάνοντας και τον άγνωστο τύπο, αποτελούνται

συνολικά από 19 χρήστες (10,1%) και μάλιστα για τους 10 χρήστες είναι άγνωστα τα χαρακτηριστικά τους.

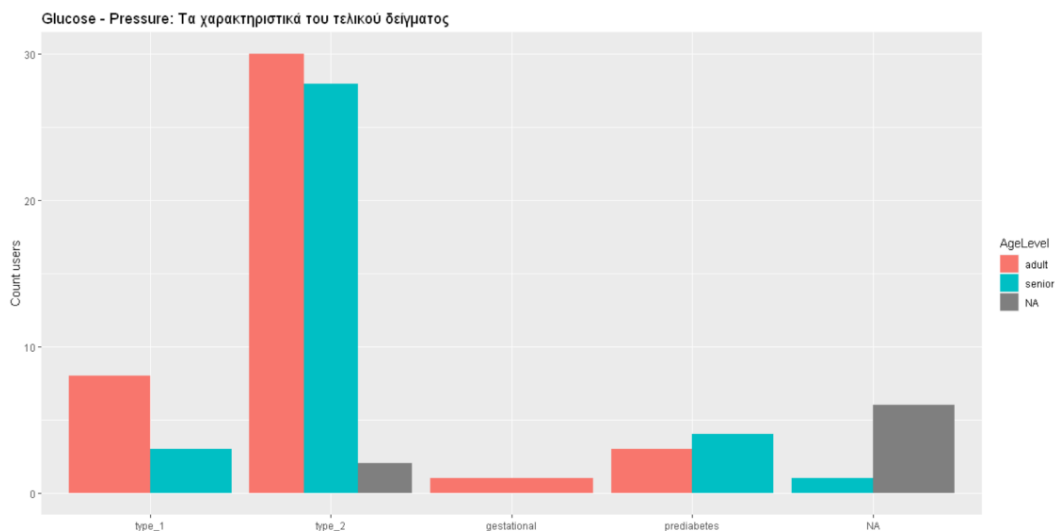
4.3.4 Μετρήσεις Pressure – Glucose

Ο σακχαρώδες διαβήτης και η αρτηριακή πίεση συνδέονται αφού η συχνότητα εμφάνισης υπέρτασης είναι διπλάσια στους διαβητικούς συγκριτικά με τον γενικό πληθυσμό. Για τον εντοπισμό του στατιστικού δείγματος που είναι συνεπείς τόσο στη μέτρηση pressure όσο και στη μέτρηση glucose, πραγματοποιείται συγχώνευση των στατιστικών δειγμάτων που προέκυψαν από την ενότητα 4.2.8 και 4.2.5. Το αποτέλεσμα απεικονίζεται στο γράφημα της Εικόνας 4.3.4.1.



Εικόνα 4.3.4.1 Οι συνεπείς χρήστες στις μετρήσεις Pressure – Glucose.

Σύμφωνα με το γράφημα της Εικόνας 4.3.4.1 οι κοινói συνεπείς χρήστες στις μετρήσεις pressure και glucose είναι 86. Συγκεκριμένα το 83,49% του δείγματος της μέτρησης pressure είναι συνεπής και στη μέτρηση glucose. Το στατιστικό δείγμα των κοινών συνεπών χρηστών παρουσιάζουν τα χαρακτηριστικά που απεικονίζονται στο γράφημα της εικόνας 4.3.4.2.

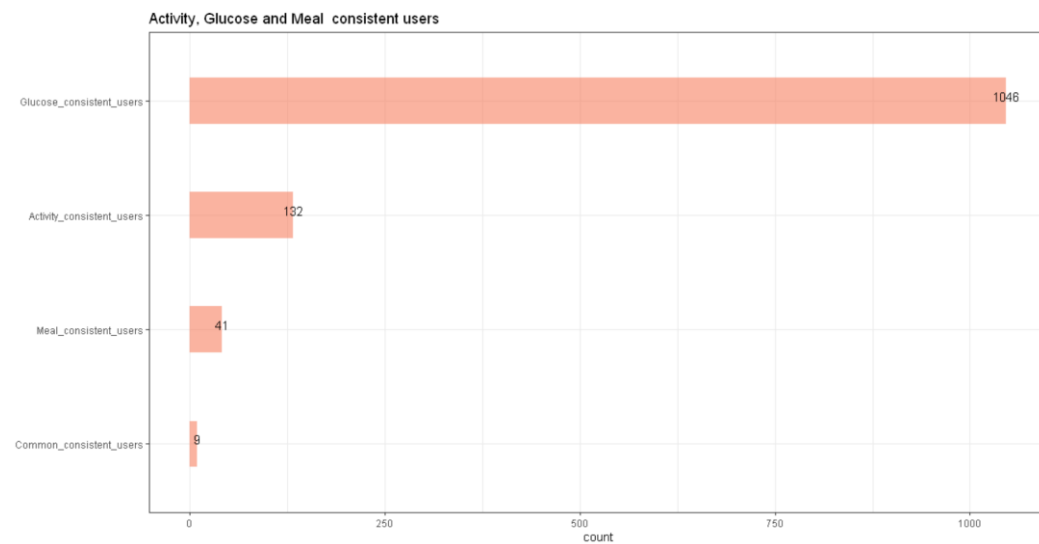


Εικόνα 4.3.4.2 Τα χαρακτηριστικά των κοινών συνεπών χρηστών στις μετρήσεις Pressure και Glucose.

Η πρώτη παρατήρηση που προκύπτει από το γράφημα της εικόνας 4.3.4.2, που παρουσιάζει τα χαρακτηριστικά του στατιστικού δείγματος που είναι συνεπείς στις μετρήσεις pressure και glucose, είναι ότι το δείγμα αποτελείται μόνο από ενήλικες (48,8%) και υπερήλικες (41,9%) χρήστες. Υπάρχουν 8 χρήστες για τους οποίους είναι άγνωστη η ηλικιακή ομάδα. Όσον αφορά τον ΣΔ τύπο του στατιστικού δείγματος, οι πλειοψηφία των χρηστών είναι διαβητικοί ΣΔ τύπου 2 (69,8%) και ακολουθούν οι χρήστες ΣΔ τύπου 1(12,8%). Οι υπόλοιποι ΣΔ τύποι έχουν ελάχιστο αριθμό χρηστών αποτελούνται συνολικά από 15 χρήστες και συνιστούν ποσοστό 17,44% των κοινών συνεπών χρηστών.

4.3.5 Μετρήσεις Activity – Glucose – Meal

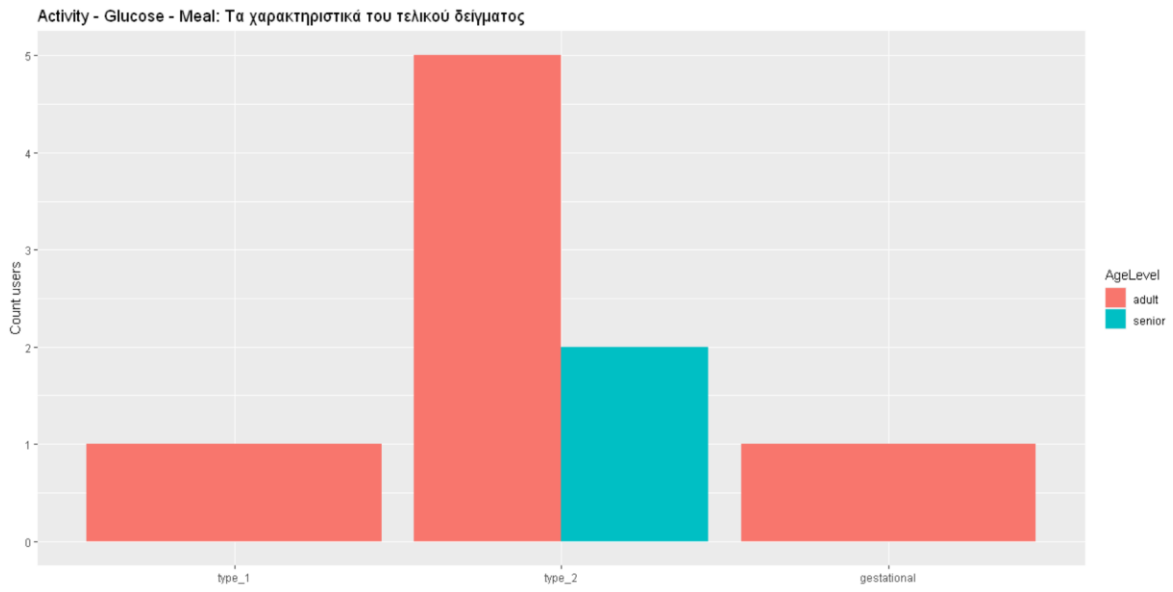
Για τον εντοπισμό των χρηστών που είναι συνεπείς στη μέτρηση activity, glucose και meal, έγινε συγχώνευση των κοινών συνεπών χρηστών από τις μετρήσεις Activity – Glucose (ενότητα 4.3.1) και τους αντίστοιχους από τις μετρήσεις Meal-Glucose (ενότητα 4.3.2). Το αποτέλεσμα φαίνεται στο γράφημα 4.3.5.1.



Εικόνα 4.3.5.1 Οι συνεπείς χρήστες στις μετρήσεις Activity – Glucose – Meal.

Σύμφωνα με το γράφημα της Εικόνας 4.3.3.1, οι κοινί συνεπείς χρήστες στις μετρήσεις activity, glucose και meal είναι μόνο 9 χρήστες. Συγκεκριμένα το 21,95% των συνεπών χρηστών της μέτρησης meal είναι συνεπής στη μέτρηση glucose και activity. Το 6,8% των συνεπών χρηστών της μέτρησης activity είναι συνεπής στην μέτρηση glucose και meal.

Στο γράφημα της εικόνας 4.3.5.2 απεικονίζονται τα χαρακτηριστικά του στατιστικού δείγματος που είναι συνεπείς στους τρεις τύπους μετρήσεων. Η πρώτη παρατήρηση που προκύπτει από το γράφημα της Εικόνας 4.3.5.2, που αντιστοιχεί στα χαρακτηριστικά του στατιστικού δείγματος που είναι συνεπείς στους τρεις τύπους μετρήσεων, είναι ότι το δείγμα αποτελείται μόνο από ενήλικες (77,8%) και υπερήλικες(22,2%) χρήστες.. Όσον αφορά τον τύπο ΣΔ του στατιστικού δείγματος, οι πλειοψηφία των χρηστών είναι διαβητικοί ΣΔ τύπου 2 (77,8%). Τόσο ο ΣΔ τύπου 1 και ο ΣΔ κήσης αποτελούνται από ένα χρήστη.



Εικόνα 4.3.5.2 Χαρακτηριστικά των κοινών συνεπών χρηστών στις μετρήσεις Activity – Glucose –Meal.

Κεφάλαιο 5ο: ΠΡΟΤΕΙΝΟΜΕΝΕΣ ΕΠΕΚΤΑΣΕΙΣ

Σε συνέχεια της διερευνητικής ανάλυσης των δεδομένων της εφαρμογής *forDiabetes*, διαμορφώθηκαν και παρατίθενται δύο (2) προτάσεις επέκτασης/βελτίωσης της εφαρμογής. Η πρώτη πρόταση αφορά στην παρουσίαση σε ένα ενιαίο γράφημα της μεταβολής των τριών βασικών μετρήσεων που καταχωρεί ο χρήστης στην τρέχουσα έκδοση της εφαρμογής: της ποσότητας γλυκόζης στο αίμα, της καύσης θερμίδων με την άσκηση, και της κατανάλωσης υδατανθράκων. Χρησιμοποιήθηκαν τα δεδομένα της *forDiabetes* για την κατασκευή του εν λόγω γραφήματος 3-σε-1 ώστε να αναδειχθεί η χρησιμότητα της πληροφορίας η οποία θα επικοινωνείται στον χρήστη. Τεκμηριώνεται επίσης η αναγκαιότητα εισαγωγής δεδομένων τα οποία να αφορούν στην ποιότητα του ύπνου του χρήστη. Θεωρήθηκε ότι οι προτεινόμενες επεκτάσεις/βελτιώσεις θα προκαλέσουν το ενδιαφέρον των χρηστών και, με αυτόν τον τρόπο, θα αυξήσουν το κίνητρό τους ώστε να χρησιμοποιούν την εφαρμογή πιο συστηματικά, καταχωρώντας δεδομένα σε καθημερινή βάση και να ελέγχουν με τον πλέον παραστατικό τρόπο την εξέλιξη των μετρήσεών τους, σε καθημερινή βάση. Με την άποψη αυτή συντάχθηκαν χωρίς καμία επιφύλαξη τα μέλη της ομάδας ανάπτυξης της εφαρμογής στην εταιρεία Tesser Multimedia, όταν τους έγινε η σχετική παρουσίαση.

5.1 Ενιαίο Γράφημα 3-σε-1

Στην περίπτωση της νόσου του Σακχαρώδους Διαβήτη, είναι πολύ ενδιαφέρον και σημαντικό για έναν διαβητικό η συσχέτιση των μετρήσεων που σχετίζονται με το επίπεδο της γλυκόζης στο αίμα και η οπτικοποίηση τους σε ένα ενιαίο γράφημα. Πιο συγκεκριμένα, μετρήσεις που αφορούν στην σωματική άσκηση, στη μέτρηση της γλυκόζης και στην κατανάλωση υδατανθράκων.

Στο πλαίσιο του σχεδιασμού του αλγόριθμου αναγωγής των τριών μετρήσεων του χρήστη σε ένα ενιαίο γράφημα, ιδανικό θα ήταν οι χρήστες να ομαδοποιηθούν ως προς τα χαρακτηριστικά τους που κρίνονται να είναι ενδιαφέροντα και αφορούν στη πάθησή τους. Για παράδειγμα, το είδος του σακχαρώδους διαβήτη από τον οποίο πάσχουν, η ηλικία, το βάρος, το φύλο, τα τυχόν επιπλέον υποκείμενα νοσήματά τους, κλπ. Για την κάθε μία ομάδα/κατηγορία χρηστών, οι ειδικοί ιατροί και διαιτολόγοι πιθανόν να είναι σε θέση να καθορίσουν τις ενδεδειγμένες τιμές ημερήσιας κατανάλωσης υδατανθράκων, άσκησης και γλυκόζης στο αίμα του διαβητικού.

Στην περίπτωση των δεδομένων της εφαρμογής *forDiabetes* τα παραπάνω δεν είναι εφικτό να γίνουν για δύο (2) κυρίως λόγους:

(α) η συλλογή των δεδομένων της εφαρμογής δεν καταχωρεί ικανό αριθμό συνεπών χρηστών οι οποίοι καταχωρούν συστηματικά και για ικανό συνεχές χρονικό διάστημα μετρήσεις. Αυτό σημαίνει ότι η συγκεκριμένη βάση δεδομένων δεν επαρκεί ώστε να συμπεριλαμβάνει την απαιτούμενη «κρίσιμη μάζα» για την κάθε μία κατηγορία χρηστών για να υπολογιστούν τα αντίστοιχα στατιστικά μεγέθη προς αναζήτηση και καθορισμό των βέλτιστων μετρήσεων ανά κατηγορία, και

(β) η συνεργασία με ικανό αριθμό ιατρών και διαιτολόγων για τον καθορισμό των βέλτιστων ημερήσιων τιμών στις μετρήσεις ανά κατηγορία χρήστη της εφαρμογής κείται εκτός των στόχων και των δυνατοτήτων της παρούσας πτυχιακή εργασίας.

Λαμβάνοντας υπόψη τα παραπάνω, υιοθετήθηκε μία απλούστερη προσέγγιση για την επίτευξη του στόχου της αναγωγής των τριών τύπων μετρήσεων σε ένα ενιαίο γράφημα: οι (ανά τύπο μέτρησης) μετρήσεις του κάθε ενός χρήστη να βαθμολογούνται και να ανάγονται σε μία κλίμακα ποιότητας στο διάστημα [0,...,100] σε σχέση με τις προγενέστερες καταχωρημένες τιμές του ίδιου τύπου μέτρησης

του συγκεκριμένου χρήστη. Το σχήμα είναι ανάλογο περίπτωσης αθλητή ο οποίος αξιολογείται σε σχέση με τις προγενέστερες επιδόσεις του. «Χρυσό» μετάλλιο (δηλ. βαθμός «100») για τον συγκεκριμένο αθλητή αποτελεί η επίτευξη επίδοσης καλύτερης από όλες τις έως τώρα επιδόσεις του στο άθλημα. Στο ενιαίο γράφημα 3-σε-1, οι μετρήσεις της σωματικής άσκησης και οι υδατάνθρακες που καταναλώθηκαν ανάγονται σε ισοδύναμη, κοινή, κλίμακα μέτρησης. Οι τιμές της τελευταίας απεικονίζονται στο διάστημα $[0, \dots, 100]$. Με εξαίρεση την μέτρηση της γλυκόζης.

Στην περίπτωση της μέτρησης της γλυκόζης, η καλύτερη επίδοση συνιστά κατά κανόνα η μικρή μέση ημερήσια τιμή γλυκόζης, αρκεί η τελευταία να μην κυμαίνεται σε τιμές μικρότερες του κατώτατου επιτρεπτού ορίου. Οι μετρήσεις της γλυκόζης πρέπει να κυμαίνονται μεταξύ 80 και 120. Εξαιτίας αυτής της ιδιαιτερότητας οι μετρήσεις της γλυκόζης αφήνονται στην απευθείας τιμή τους στο ενιαίο γράφημα 3-σε-1, χωρίς να συγκρίνονται με τις παρελθούσες μετρήσεις του ίδιου χρήστη. Στην περίπτωση σύγκρισης με τις προγενέστερες καταχωρημένες τιμές, θα βαθμολογηθεί με 100 η μικρότερη τιμή γλυκόζης του χρήστη, που μπορεί να σημαίνει και υπογλυκαιμία (τιμή γλυκόζης < 80).

Προς αυτήν την κατεύθυνση, ένα πρώτο βήμα για κάθε τύπο μέτρησης δημιουργήθηκε data frame με τα πεδία «ημερομηνία» - «τιμή». Όπου το πεδίο «ημερομηνία», είναι η ημερομηνία της καταχώρησης του χρήστη και «τιμή» είναι η τιμή που καταχώρησε στον αντίστοιχο τύπο μέτρησης. Ο χρήστης έχει τη δυνατότητα να καταχωρεί περισσότερες από μια φορά τιμές στους παραπάνω τύπου μέτρησης. Για τον λόγο αυτό εφαρμόστηκε ο αλγόριθμος που ομαδοποιεί (ανά τύπο μέτρησης) μετρήσεις ανά ημέρα για τον κάθε χρήστη:

- Για κάθε τύπο μέτρησης
 - Για κάθε μια ημερομηνία
 - Υπολογίζεται για το σύνολο των καταχωρήσεων που έγιναν εκείνη την ημερομηνία:
 - Για τη σωματική άσκηση, το σύνολο των θερμίδων που κάηκαν σε κάθε δραστηριότητα
 - Για την γλυκόζη, ο μέσος όρος των τιμών της γλυκόζης που μετρήθηκαν
 - Για τα γεύματα, το σύνολο των υδατανθράκων που καταναλώθηκαν σε κάθε γεύμα
 - Τα αποτελέσματα αποθηκεύτηκαν σε ένα data frame που καταχωρεί ζεύγη τιμών «ημερομηνία»-«αποτέλεσμα». Παράδειγμα αποτελεί ο Πίνακας 5.2.1.

Στη συνέχεια, στην επόμενη ενότητα εφαρμόζεται ο τύπος Rainbow Ranking για αυτή κάθε αυτή την αναγωγή σε κοινή κλίμακα των μετρήσεων που αφορούν τις θερμίδες που κάηκαν στη σωματική άσκηση και τους υδατάνθρακες που καταναλώθηκαν στα γεύματα.

Date	Calories
2019-03-21	15275
2019-04-04	11531
2019-04-10	10961
2019-03-28	10278
2019-04-08	7891
2019-03-27	7058
2019-04-03	4355
2019-07-26	4218
2019-09-11	3640
2019-03-23	3429
2019-03-07	3408
2019-03-08	3272
2019-04-02	3212

Πίνακας 5.1.1. Στιγμιότυπο data frame ανά χρήστη μετά την εκτέλεση του αλγόριθμου. Το συγκεκριμένο, αφορά τις θερμίδες που έκαψε ο χρήστης συνολικά λόγω σωματικής άσκησης στην αντίστοιχη ημερομηνία.

5.1.1 Κοινή κλίμακα μέτρησης

Ανάλογα με τον τύπο της μέτρησης, διαφορετικές είναι οι μονάδες που χρησιμοποιούνται. Μονάδα μέτρησης για τις θερμίδες είναι το kcal και για τους υδατάνθρακες τα γραμμάρια (gr). Για τον κάθε ένα χρήστη και για τον κάθε ένα τύπο μέτρησης, οι εγγραφές (ανάλογες εκείνων του Πίνακα 5.1.1) ταξινομούνται κατά φθίνουσα τάξη. Για το λόγο αυτό μετασχηματίζονται και γίνονται σε αναγωγή οι τιμές της μέτρησης που αφορούν τη σωματική άσκηση και τη κατανάλωση υδατανθράκων σε μια κοινή κλίμακα στο διάστημα $[0, \dots, 100]$ με υπολογισμό της τιμής Rainbow Ranking [25].

Πιο συγκεκριμένα, για κάθε τιμή c του data frame για τον x τύπο μέτρησης (σωματική άσκηση, κατανάλωση υδατανθράκων) γίνεται χρήση του τύπου Rainbow Ranking για το μετασχηματισμό των τιμών σε τιμές $RR_{c,x}$ στο διάστημα $[0, 100]$:

$$RR_{c,x} = 100 - 100 * \left(\frac{|N_{above}(c)|}{|C|} + \frac{|N_{tie}(c)|}{2|C|} \right)$$

όπου:

- $RR_{c,x}$ είναι η τιμή RR για τον τύπο μέτρησης x ,
- C ο αριθμός των εγγραφών στο συγκεκριμένο data frame
- $N_{above(c)}$ ο αριθμός των παρατηρήσεων που ταξινομούνται να είναι μεγαλύτερα της τιμής c
- $N_{tie(c)}$ ο αριθμός των παρατηρήσεων που ταξινομούνται να ισοβαθμούν με την τιμή c

Για την αναγωγή των τιμών της μέτρησης της σωματικής άσκησης και των υδατανθράκων που καταναλώθηκαν σε κοινή κλίμακα εκτελέστηκε ο εξής αλγόριθμος:

- Για τον κάθε τύπο μέτρησης data frame, της μορφής Πίνακα 5.1.1
 - Ταξινομούνται οι τιμές σε φθίνουσα σειρά

- Για κάθε τιμή
 - Εφαρμόζεται ο τύπος Rainbow Ranking για την αναγωγή των τιμών στο διάστημα [0,100]
 - Υπολογίζεται η τελική τιμή επίδοσης στη κοινή κλίμακα στο διάστημα [0,100]

Για τον κάθε τύπο μέτρησης το 100 έχει διαφορετική σημασία:

- Για την σωματική άσκηση, είναι η καλύτερη επίδοση συνιστά η καύση μεγάλου αριθμού θερμίδων.
- Για τα γεύματα, καλύτερη επίδοση συνιστά η μικρή μέση ημερήσια κατανάλωση υδατανθράκων.

Date	Calories	perRank	perRankrr
2019-03-21	15275	1.0000000	100.0
2019-04-04	11531	0.9966102	99.7
2019-04-10	10961	0.9932203	99.3
2019-03-28	10278	0.9898305	99.0
2019-04-08	7891	0.9864407	98.6
2019-03-27	7058	0.9830508	98.3
2019-04-03	4355	0.9796610	98.0
2019-07-26	4218	0.9762712	97.6
2019-09-11	3640	0.9728814	97.3
2019-03-23	3429	0.9694915	96.9
2019-03-07	3408	0.9661017	96.6
2019-03-08	3272	0.9627119	96.3
2019-04-02	3212	0.9593220	95.9
2019-04-07	3040	0.9559322	95.6
2019-03-05	2979	0.9525424	95.3
2019-09-03	2973	0.9491525	94.9
2019-08-27	2788	0.9457627	94.6
2019-09-04	2719	0.9423729	94.2

Πίνακας 5.1.1.1 Στιγμιότυπο από το αποτέλεσμα του υπολογισμού τιμών RR (περίπτωση: calories).

Ο Πίνακας 5.1.1.1 αντιστοιχεί στο αποτέλεσμα που παράγει η εκτέλεση του αλγορίθμου αναγωγής σε κοινή κλίμακα για τη μέτρηση της σωματικής άσκησης. Είναι ένας πίνακας που καταχωρεί τις προαναφερθείσες τιμές, ταξινομημένες κατά φθίνοντα αναγόμενη τιμή. Το αντίστοιχο αποτέλεσμα έχουν και οι άλλοι τύποι μέτρησης.

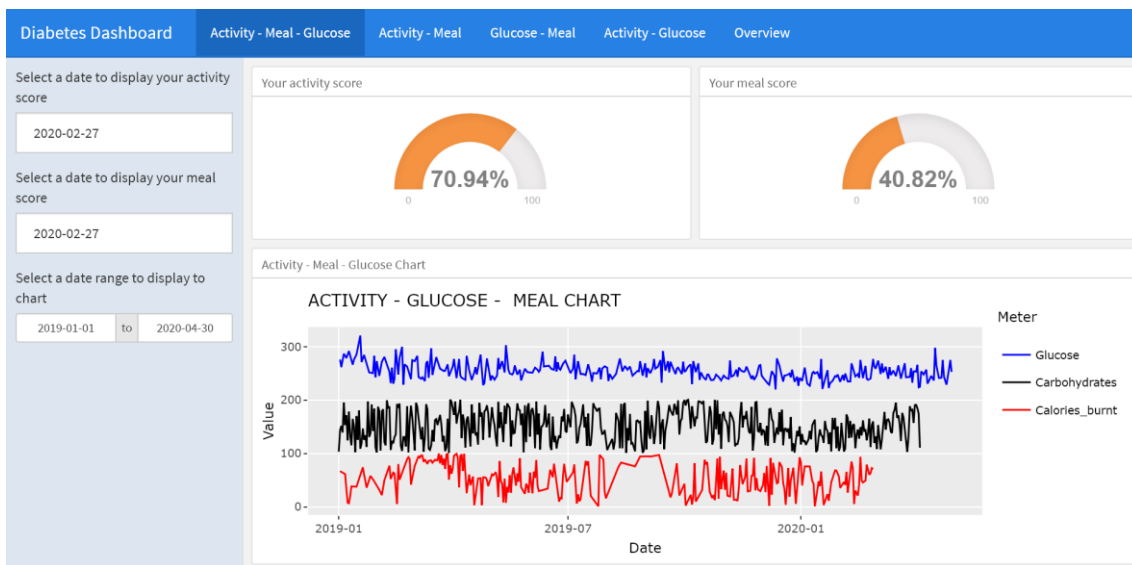
5.1.2 Οπτικοποίηση του Γραφήματος 3-σε-1

Για την οπτικοποίηση της συσχέτισης των τριών μετρήσεων δημιουργήθηκε γράφημα 3-σε-1. Στο άξονα x αναπαριστάται η ημερομηνία της καταχώρησης, στο άξονα y η τιμή της αντίστοιχης μέτρησης. Πιο αναλυτικά στην κλίμακα στον άξονα y, από το 0-100 είναι η επίδοση του χρήστη στη σωματική άσκηση, από το 101-202 είναι η επίδοση του χρήστη στη κατανάλωση υδατανθράκων και από το 203 μέχρι όσο πάει η μέση τιμή της γλυκόζης (το 203 αντιστοιχεί στη μέση τιμή γλυκόζης 60). Κάθε καμπύλη αναπαριστά την πορεία της επίδοσης του χρήστη για τον κάθε ένα τύπο μέτρησης.

5.1.3 Χρηστική Διεπαφή

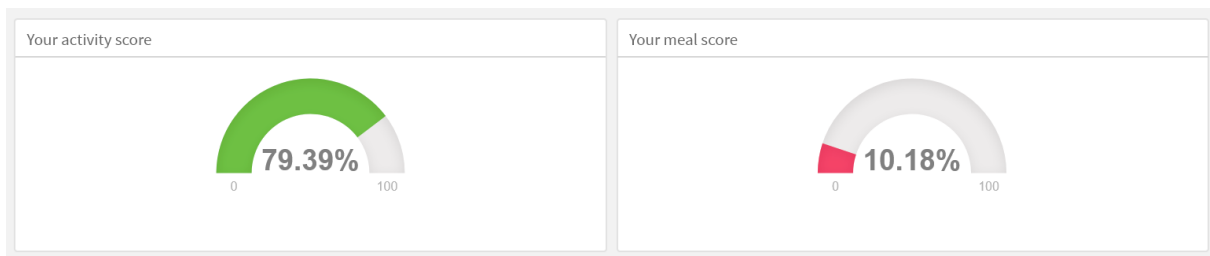
Για την παρουσίαση των παραπάνω στοιχείων, δημιουργήθηκε μια χρηστική διεπαφή. Με τη χρήση των βιβλιοθηκών shiny¹ και flexdashboard², το προγραμματιστικό περιβάλλον της R παρέχει τη δυνατότητα δημιουργίας διαδραστικών γραφημάτων. Επιτρέπει στο χρήστη να αλλάζει τις παραμέτρους στα εικονιζόμενα στοιχεία και να παρακολουθεί πως μεταβάλλονται οι γραφικές απεικονίσεις.

Παρατηρώντας την Εικόνα 5.1.3.1, που αντιστοιχεί στη χρηστική διεπαφή, χωρίζεται σε τρία μέρη. Το πάνω μέρος (μπλε περιοχή) είναι το μενού, που περιέχει τις καρτέλες από τις οποίες αποτελείται η χρηστική διεπαφή. Στα αριστερά είναι τα πεδία στα οποία ο χρήστης εισάγει τιμές, στη συγκεκριμένη διεπαφή ημερομηνίες. Στα δεξιά απεικονίζονται τα γραφήματα, τα οποία μεταβάλλονται ανάλογα την ημερομηνία που εισάγει ο χρήστης.



Εικόνα 5.1.3.1 Γράφημα 3-σε-1.

Ακολουθεί αναλυτική περιγραφή της καρτέλας Activity – Glucose – Meal που απεικονίζεται στην εικόνα 5.1.3.1. Περιέχει δυο πλαίσια στα οποία απεικονίζεται η επίδοση του χρήστη στις μετρήσεις της σωματικής άσκησης και κατανάλωσης υδατανθράκων. Ακόμη, περιλαμβάνει ένα πλαίσιο με το γράφημα 3-σε-1.



Εικόνα 5.1.3.3 Γραφική απεικόνιση της επίδοσης του χρήστη.

Σύμφωνα με την Εικόνα 5.1.3.2 για την εμφάνιση της επίδοσης για τη μέτρηση της σωματικής άσκησης και κατανάλωσης υδατανθράκων, υπάρχουν δύο πλαίσια που περιέχουν ένα μετρητή/κοντέρ.

¹ <https://shiny.rstudio.com/>

² <https://rmarkdown.rstudio.com/flexdashboard/>

Ο χρήστης εισάγει την ημερομηνία που επιθυμεί και εμφανίζεται στην αντίστοιχη περιοχή η επί τις % τιμή επίδοσης του (score) που είχε τη συγκεκριμένη ημερομηνία.



Εικόνα 5.1.3.2 Γράφημα 3-σε-1: καθορισμός ημερομηνίας για τον κάθε έναν τύπο μέτρησης, χωριστά.

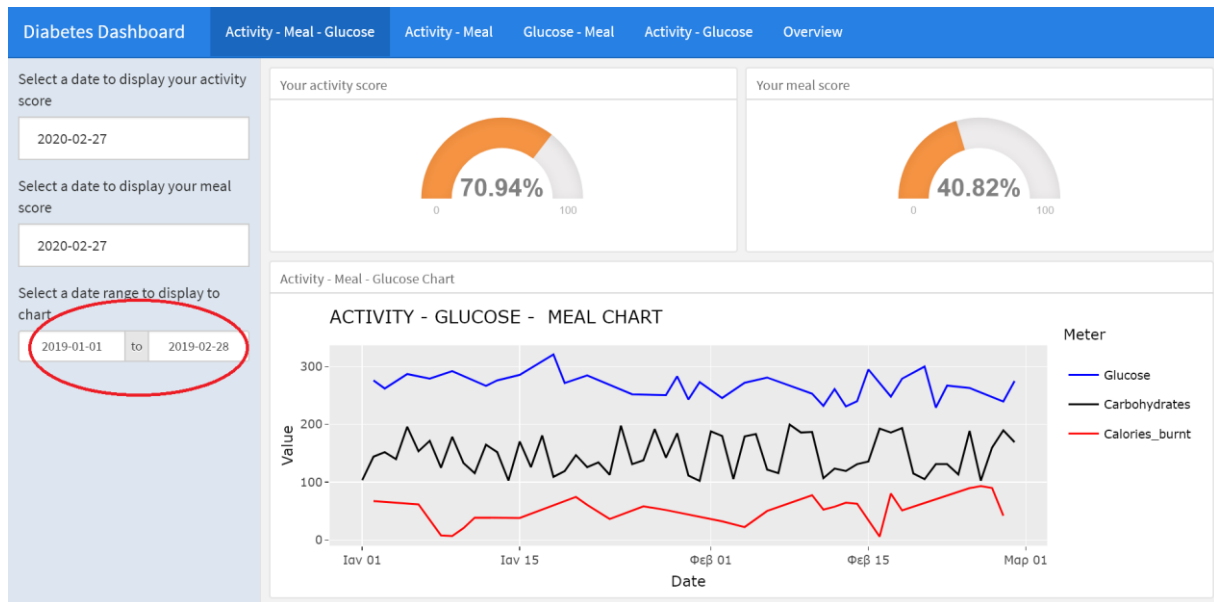
Ο συγκεκριμένος τύπος γραφικής απεικόνισης της επίδοσης ονομάζεται gauge, εμφανίζει μια αριθμητική τιμή σε έναν μετρητή που κυμαίνεται μεταξύ καθορισμένων ελάχιστων και μέγιστων τιμών. Αυτός ο τύπος γραφήματος βοηθάει στην καλύτερη κατανόηση της επίδοσης του χρήστη, γιατί αν η επίδοση είναι πολύ χαμηλή στο διάστημα [0,39] απεικονίζεται με κόκκινο χρώμα, αν είναι μέτρια στο διάστημα [40,74] απεικονίζεται με πορτοκαλή και πάνω από το όριο στο διάστημα [75,100] απεικονίζεται με πράσινο χρώμα. Παράδειγμα αποτελεί η Εικόνα 5.1.3.3.

Επιπλέον, ο χρήστης στη διεπαφή όπως φαίνεται και στην Εικόνα 5.1.3.2 έχει τη δυνατότητα να βλέπει την πορεία της επίδοσης του σε σχέση με τους άλλους τύπους μέτρησης αλλά και όλο το χρονικό διάστημα που ήταν ενεργός (από την πρώτη μέχρι την τελευταία καταχώρηση του) στην εφαρμογή σε ένα ενιαίο γράφημα. Κάθε χρώμα αναπαριστά ένα διαφορετικό τύπο μέτρησης. Η μπλε καμπύλη Glucose αναπαριστά τις μετρήσεις της γλυκόζης (εμφανίζονται στη κλίμακα στο άξονα y από το 203 μέχρι όπου φτάσει). Η μαύρη καμπύλη Carbohydrates αναπαριστά τις μετρήσεις των υδατανθράκων που καταναλώθηκαν (εμφανίζονται στη κλίμακα στο άξονα y από το 101-202). Η κόκκινη καμπύλη Calories_burnt αναπαριστά τις μετρήσεις των θερμίδων που κάηκαν στη σωματική άσκηση (εμφανίζονται στη κλίμακα στο άξονα y από το 0-101). Για την καλύτερη κατανόηση και ερμηνεία του γραφήματος, η καμπύλη Carbohydrates στο γράφημα αναπαριστά την τιμή της επίδοσης του χρήστη ανάποδα (κάνει κορυφές στο γράφημα όταν έχει καταναλώσει ο χρήστης τις περισσότερες θερμίδες). Επιπλέον, ο χρήστης μπορεί να επιλέξει το χρονικό διάστημα που τον ενδιαφέρει να απεικονιστεί στο γράφημα.

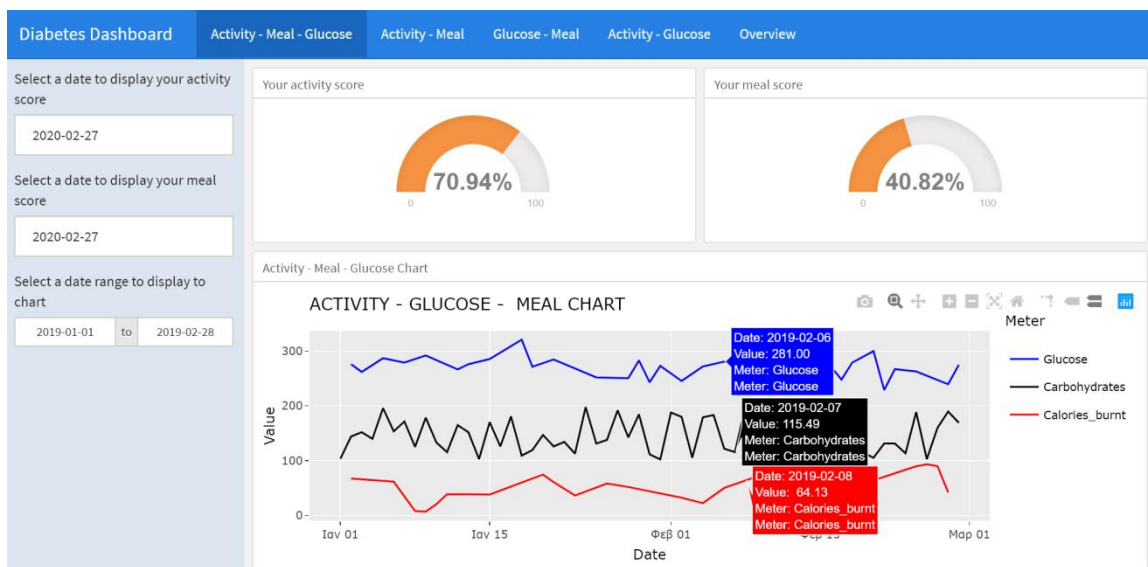
Η Εικόνα 5.1.3.4 αντιστοιχεί στο αποτέλεσμα του γραφήματος 3-σε-1 που εμφανίζεται στο χρήστη μετά τον καθορισμό του χρονικού διαστήματος(2019-01-01 έως 2019-02-28). Στο γράφημα ο χρήστης μπορεί να παρατηρήσει ότι:

- Στο διάστημα 1/1 έως 15/1, όταν δεν κάνει έντονη άσκηση και καταναλώνει αρκετούς υδατάνθρακες η τιμή της γλυκόζης του ανεβαίνει.

- Στο διάστημα 15/1 έως 1/2, όταν ανεβαίνει η επίδοση του στη γυμναστική και ελαττώνεται η επίδοση του στο γεύμα, η τιμή της γλυκόζης ελαττώνεται με μια μικρή καθυστέρηση.
- Στο διάστημα 15/2 έως 28/2, στις 16/2 είναι η χειρότερη επίδοση στη άσκηση και καταναλώνει τους περισσότερους υδατάνθρακες, η τιμή της γλυκόζης είναι στην κορυφή
- Στο ίδιο διάστημα, στο σημείο που κοντεύει να ενωθούν οι καμπύλες της άσκησης και του γεύματος, η τιμή της γλυκόζης αρχίζει να ελαττώνεται.
- Στα τέλη Φεβρουαρίου, όταν ελαττώθηκε η γλυκόζη, σταμάτησε την άσκηση και άρχισε να καταναλώνει πολλούς υδατάνθρακες, οπότε ξεκίνησε η άνοδος της καμπύλης γλυκόζης.



Εικόνα 5.1.3.4 Γράφημα 3-σε-1: μετά τον καθορισμό του χρονικού διαστήματος.

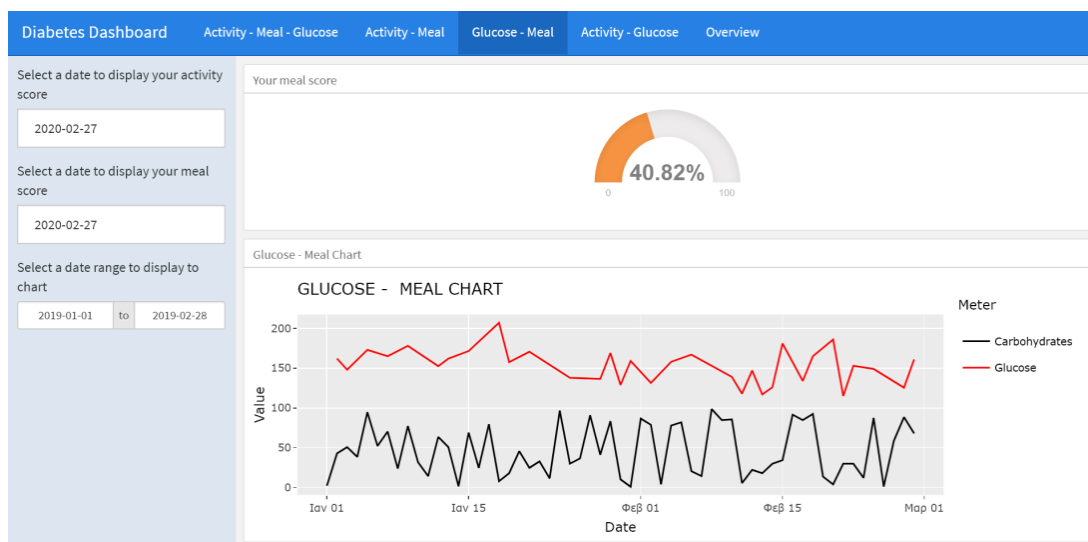


Εικόνα 5.1.3.5 Γράφημα 3-σε-1: εμφάνιση επιπλέον πληροφορίας

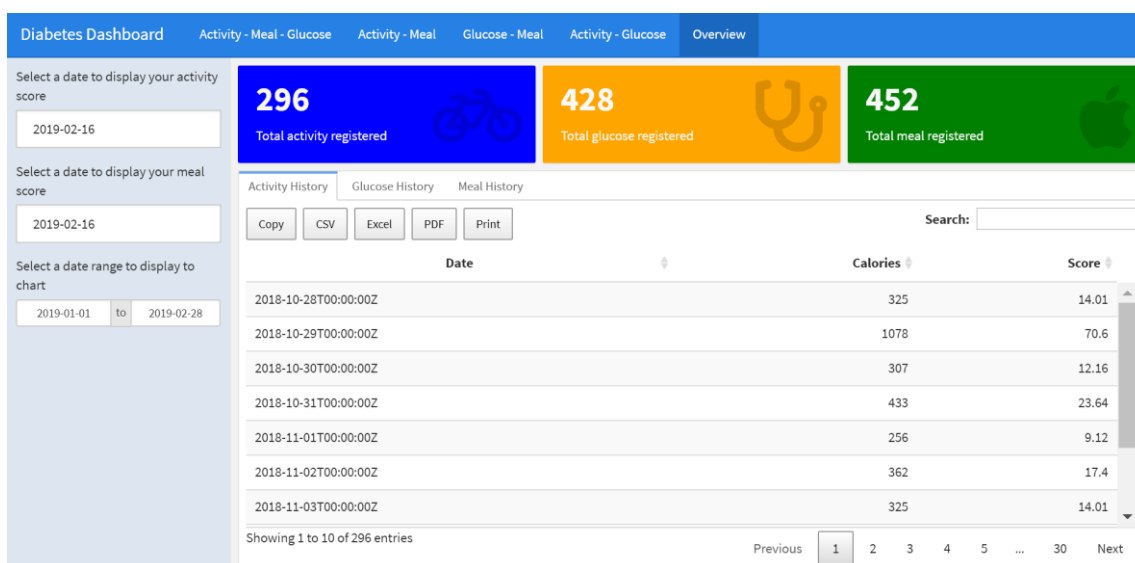
Ακόμη, η χρηστική διεπαφή παρέχει στον χρήστη επιπλέον εργαλεία/δυνατότητες για το γράφημα 3-σε-1. Η Εικόνα 5.1.3.5 απεικονίζει τις δυνατότητες που έχει ο χρήστης πάνω στο γράφημα 3-σε-1.

Μπορεί να βλέπει και να συγκρίνει για όποιο σημείο των ενδιαφέρει την ημερομηνία και την τιμή της μέτρησης. Επιπλέον, μπορεί να αποθηκεύσει ως εικόνα το γράφημα, να κάνει zoom in και zoom out.

Οι ίδιες ακριβώς λειτουργίες και δυνατότητες εκτελούνται και στις άλλες τρεις καρτέλες. Με μόνη διαφορά ότι συσχετίζονται τα ζευγάρια μετρήσεων που αναγράφονται στο όνομα της καρτέλας. Για παράδειγμα στην καρτέλα Glucose – Meal παρουσιάζεται το ενιαίο γράφημα με τις καμπύλες της κατανάλωσης υδατανθράκων και τις τιμές της γλυκόζης.



Εικόνα 5.1.3.6 Καρτέλα Glucose-Meal στη χρηστική διεπαφή.



Εικόνα 5.1.3.7 Στιγμιότυπο καρτέλας Overview στη χρηστική διεπαφή.

Στο μενού στη χρηστική διεπαφή υπάρχει η καρτέλα Overview. Όπως φαίνεται και στην Εικόνα 5.1.3.7 που αντιστοιχεί στο στιγμιότυπο της καρτέλας Overview υπάρχουν τρία χρωματικά πλαίσια. Το μπλε πλαίσιο περιέχει το σύνολο των ημερών που ο χρήστης έχει καταχωρήσει τιμές στη σωματική άσκηση, αντίστοιχα το κίτρινο για την γλυκόζη και το πράσινο για τα γεύματα.

Επιπλέον, η καρτέλα Overview περιέχει για κάθε τύπο μέτρησης έναν πίνακα με το ιστορικό των τιμών και των επιδόσεων του χρήστη. Όπως απεικονίζεται στην Εικόνα 5.1.3.8, ο χρήστης μπορεί να κάνει αναζήτηση στον πίνακα με οποιαδήποτε τιμή/πεδίο. Στην εικόνα 5.1.3.8, γίνεται αναζήτηση,

στο πίνακα με τις μετρήσεις της σωματικής άσκησης, βάσει της ημερομηνίας καταχώρησης. Στη συνέχεια ο χρήστης μπορεί να το αποθηκεύσει το αποτέλεσμα ή ολόκληρο τον πίνακα σε αρχείο csv, excel και pdf, επιλέγοντας το αντίστοιχο κουμπί/ πλαίσιο.

The screenshot shows a dashboard with three main sections: Activity History (296 registered), Glucose History (428 registered), and Meal History (452 registered). Below these are tabs for 'Activity History', 'Glucose History', and 'Meal History'. A search bar is highlighted with a red circle, containing the text 'Search: 2019-01'. Below the search bar are buttons for 'Copy', 'CSV', 'Excel', 'PDF', and 'Print'. A table displays activity data with columns for Date, Calories, and Score. The table shows 10 entries, with a total of 12 entries displayed (filtered from 296 total entries). Navigation buttons for 'Previous', '1', '2', and 'Next' are visible at the bottom.

Date	Calories	Score
2019-01-02T00:00:00Z	987	67.06
2019-01-06T00:00:00Z	843	61.48
2019-01-08T00:00:00Z	243	7.77
2019-01-09T00:00:00Z	221	6.75
2019-01-10T00:00:00Z	393	20.27
2019-01-11T00:00:00Z	566	38.51
2019-01-15T00:00:00Z	553	37.83

Εικόνα 5.1.3.8 Εντοπισμός μετρήσεων σε συγκεκριμένη χρονική περίοδο (2019-01).

Η Εικόνα 5.1.3.9 αντιστοιχεί στο αρχείο pdf που αποθηκεύεται από τον πίνακα της καρτέλας Overview. Στο έγγραφο pdf της εικόνας 5.1.3.9, στην πρώτη στήλη είναι η ημερομηνία της καταχώρησης του χρήστη, στη δεύτερη στήλη είναι το σύνολο των θερμίδων που έκαψε στη σωματική άσκηση τη συγκεκριμένη ημερομηνία και στην τελευταία στήλη είναι η τιμή που αντιστοιχεί στην επίδοση του.

The screenshot shows a PDF document titled 'Diabetes Dashboard'. The document contains a table with three columns: Date, Calories, and Rcalories. The table lists 12 entries of activity data for January 2019. The PDF is displayed in a viewer window with a toolbar at the top.

Date	Calories	Rcalories
2019-01-02T00:00:00Z	987	67.06
2019-01-06T00:00:00Z	843	61.48
2019-01-08T00:00:00Z	243	7.77
2019-01-09T00:00:00Z	221	6.75
2019-01-10T00:00:00Z	393	20.27
2019-01-11T00:00:00Z	566	38.51
2019-01-15T00:00:00Z	553	37.83
2019-01-20T00:00:00Z	1165	74.32
2019-01-21T00:00:00Z	836	60.48
2019-01-23T00:00:00Z	532	36.14
2019-01-26T00:00:00Z	791	58.1
2019-01-28T00:00:00Z	687	51.68

Εικόνα 5.1.3.9 Αποθήκευση του αποτελέσματος της αναζήτησης σε αρχείο pdf.

5.2 Εισαγωγή του ύπνου ως μέτρηση στην εφαρμογή

Ο ύπνος είναι μια δυναμική διαδικασία, που παίζει σημαντικό ρόλο στο μεταβολισμό της γλυκόζης. Μελέτες υποστηρίζουν ότι υπάρχει σημαντική συσχέτιση μεταξύ ανεπαρκούς σε ποσότητα και ποιότητα ύπνου και σακχαρώδους διαβήτη [26].

Ο ύπνος έχει πέντε στάδια: ταχεία κίνηση των ματιών (REM) ύπνου και στάδια 1, 2, 3, και 4 του ύπνου μη REM. Τα βαθύτερα στάδια του ύπνου μη REM, δηλαδή, στάδια 3 και 4, είναι επίσης γνωστά ως ύπνος αργού κύματος και πιστεύεται ότι είναι ο πιο ωφέλιμος στο ανθρώπινο σώμα.

Η έναρξη του ύπνου με αργό κύμα σχετίζεται προσωρινά με παροδικές μεταβολικές και ορμονικές αλλαγές, οι οποίες μπορούν να επηρεάσουν την ομοιόσταση της γλυκόζης. Ως αποτέλεσμα, στο πρώτο μέρος της νύχτας, όταν κυριαρχεί ο ύπνος με αργό κύμα, ο μεταβολισμός της γλυκόζης είναι πιο αργός. Αυτή η επίδραση αντιστρέφεται στο δεύτερο μέρος της νύχτας, όταν ο ύπνος REM ή το στάδιο 1 κυριαρχεί.

Λαμβάνοντας υπόψη αυτές τις αλλαγές στο μεταβολισμό της γλυκόζης κατά τη διάρκεια του ύπνου, δεν προκαλεί έκπληξη το γεγονός ότι ο λιγότερος ή χαμηλότερος σε ποιότητας ύπνος σε συστηματική βάση θα μπορούσε να επηρεάσει τη συνολική ομοιόσταση της γλυκόζης. Παρατίθενται οι εργαστηριακές μελέτες που ενισχύουν την παραπάνω άποψη.

Η πρώτη εργαστηριακή μελέτη για την επίδραση της στέρησης του ύπνου στον μεταβολισμό δημοσιεύτηκε το 1969 [27]. Έδειξε ότι η συνολική στέρηση ύπνου οδήγησε σε σημαντική αύξηση των επιπέδων γλυκόζης. Χρειάζεται προσοχή στην παρέκταση τέτοιων αποτελεσμάτων στις πραγματικές συνθήκες ζωής, καθώς η συνολική στέρηση ύπνου είναι ασυνήθιστη στον άνθρωπο και αναπόφευκτα ακολουθείται από την αποκατάσταση του ύπνου, με ομαλοποίηση του μεταβολισμού της γλυκόζης.

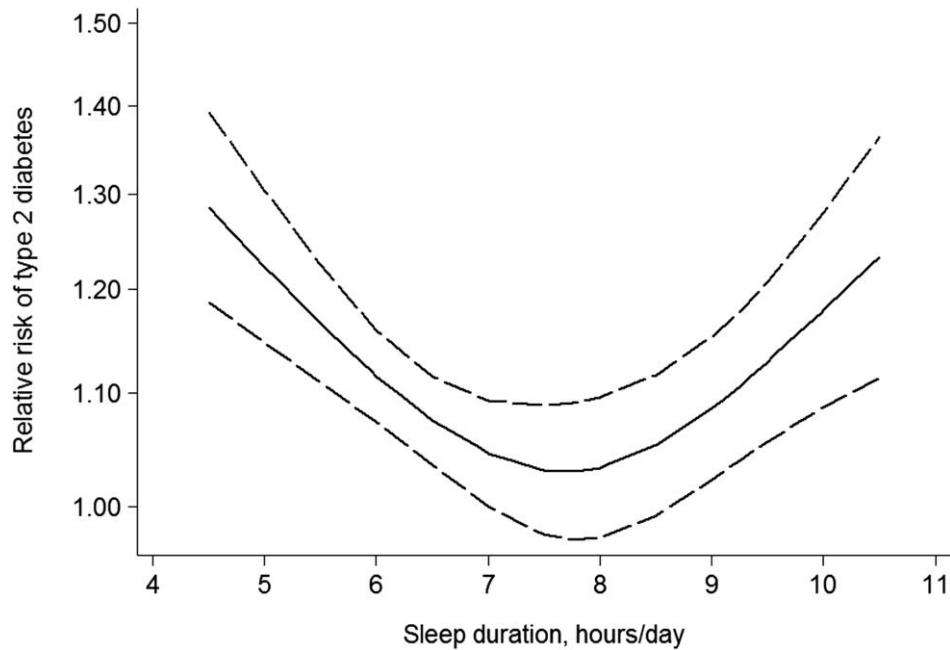
Ο Spiegel [28], σε εργαστηριακές μελέτες σχετικά με τη μερική στέρηση ύπνου σε υγιείς ενήλικες, διαπίστωσαν ότι ο περιορισμός του ύπνου σε 4 ώρες τη νύχτα για 6 νύχτες είχε ως αποτέλεσμα τη μείωση κατά 40% στην ανοχή στη γλυκόζη, σε επίπεδα παρόμοια με αυτά που παρατηρήθηκαν σε ηλικιωμένους ενήλικες με περιορισμένη ανοχή στη γλυκόζη. Οι παρατηρήσεις του Spiegel οδήγησαν σε μια σειρά επιδημιολογικών μελετών που εξέτασαν τις σχέσεις μεταξύ της διάρκειας του ύπνου και των διαταραχών του ύπνου και του κινδύνου διαβήτη.

Το Sleep Heart Study [29] ήταν μια μελέτη στην οποία αξιολόγησαν τη σχέση μεταξύ της αναφερόμενης διάρκειας ύπνου και της ανοχής σε σχέση με τη γλυκόζη ή του διαβήτη τύπου 2 σε περισσότερους από 1.400 άνδρες και γυναίκες που δεν είχαν ιστορικό αϋπνίας. Αφού διαχώρισαν το δείγμα ανάλογα την ηλικία, το φύλο, εθνικότητα, σωματική κατάσταση και το δείκτη άπνοιας-υποπνοίας, παρατηρήθηκε ότι η μειωμένη ανοχή στη γλυκόζη και προδιάθεση για διαβήτη 2 ήταν υψηλότερη σε εκείνους που ανέφεραν ύπνο 6 ώρες ή λιγότερο ανά διανυκτέρευση ή 9 ώρες ή περισσότερο ανά διανυκτέρευση.

Ο Yaggi [30], σε μια μελέτη 1.139 ανδρών στις ΗΠΑ, διαπίστωσαν επίσης μια σχέση σχήματος U μεταξύ της διάρκειας του ύπνου και της επίπτωσης του διαβήτη, με υψηλότερα ποσοστά σε άτομα που κοιμόντουσαν λιγότερο από 5 ή περισσότερες από 8 ώρες τη νύχτα. Στην Εικόνα 5.2.1 απεικονίζεται η σχέση σχήματος U μεταξύ της διάρκειας του ύπνου και του σακχαρώδους διαβήτη τύπου 2.

Ο Carpuccio [32], ανέλυσαν τις μελέτες που είχαν δημοσιευτεί μέχρι τότε που αφορούσαν τη διάρκεια του ύπνου και το σακχαρώδη διαβήτη. Υπολόγισαν ότι ο κίνδυνος εμφάνισης σακχαρώδους διαβήτη ήταν 28% υψηλότερος στα άτομα με μικρή διάρκεια ύπνου (≤ 5 ή <6 ώρες στις διάφορες

μελέτες), 48% υψηλότερος στα άτομα με μεγάλη διάρκεια ύπνου (> 8 ώρες), 57% υψηλότερος στα άτομα με δυσκολία μετάβασης για ύπνο και 84% υψηλότερο στα άτομα με δυσκολία να κοιμηθούν.



Εικόνα 5.1.1 Η σχέση μεταξύ διάρκειας ύπνου και κινδύνου διαβήτη τύπου 2 [31].

Μερικές μελέτες εξέτασαν συγκεκριμένα την επίδραση της καταστολής του ύπνου αργών κυμάτων στην ομοιόσταση γλυκόζης. Αρχικά, ο Tassali [33] εξέτασε εννέα γυναίκες και άνδρες νεαρής ηλικίας μετά από 2 συνεχόμενες νύχτες ανενόχλητου ύπνου χωρίς αλλαγή στη συνολική διάρκεια του ύπνου ή στη διάρκεια του ύπνου REM. Ο ύπνος των αργών κυμάτων διαταράσσεται από την «παροχή ακουστικών ερεθισμάτων διαφόρων συχνοτήτων και εντάσεων» όταν τα άτομα άρχισαν να πηγαίνουν στο στάδιο 3 ή στο στάδιο 4 του ύπνου. Αυτό μείωσε το ποσοστό του ύπνου αργού κύματος σχεδόν κατά 90%, το οποίο είναι συγκρίσιμο με το βαθμό κατακερματισμού του ύπνου που παρατηρείται σε μέτρια έως σοβαρή αποφρακτική άπνοια ύπνου. Μετά από 3 νύχτες καταστολής του ύπνου των αργών κυμάτων, η ινσουλίνη μειώθηκε κατά 25% και η οποία είχε ως αποτέλεσμα τη μείωση της ανοχής στη γλυκόζη κατά 23%, μια τιμή που παρατηρήθηκε σε ηλικιωμένους ενήλικες με προβλήματα ανοχής στη γλυκόζη.

Τέλος, ο Aronsohn [34] πραγματοποίησε μια μελέτη στην οποία συμμετείχαν 60 ασθενείς με σακχαρώδη διαβήτη τύπου 2. Διαπίστωσε ότι 46 (77%) από αυτούς είχαν αποφρακτική άπνοια ύπνου. Άπνοια στον ύπνο είναι κάθε διακοπή της αναπνοής διάρκειας 10 δευτερολέπτων και άνω σε κοιμώμενο άτομο. Επιπλέον, όσο χειρότερη είναι η αποφρακτική άπνοια ύπνου, τόσο χειρότερος είναι ο έλεγχος της γλυκόζης. Σε σύγκριση με ασθενείς χωρίς αποφρακτική άπνοια ύπνου, ο προσαρμοσμένος μέσος όρος της γλυκοζυλιωμένης αιμοσφαιρίνης A1c αυξήθηκε σε γραμμική τάση κατά (σε απόλυτες ποσοστιαίες μονάδες):

- 1,49% σε ασθενείς με ήπια αποφρακτική άπνοια ύπνου ($P = .0028$)
- 1,93% σε ασθενείς με μέτρια αποφρακτική άπνοια ύπνου ($P = .0033$)
- 3,69% σε ασθενείς με σοβαρή αποφρακτική άπνοια ($P < .0001$).

Κεφάλαιο 6ο: ΕΠΙΛΟΓΟΣ

6.1 Σύνοψη

Η παρούσα πτυχιακή εργασία αποτελεί παράδειγμα της συμβολής της ακαδημαϊκής έρευνας στους φορείς ανάπτυξης/παραγωγής, όπως είναι η εταιρεία Tesseta Multimedia.

Από την παραγωγή γραφημάτων κατανομής των χρηστών/ασθενών ως προς τον τύπο Σακχαρώδους Διαβήτη, την ηλικία και τη χώρα, γίνεται πιο εύκολα και γρήγορα κατανοητή η συνολική εικόνα των χρηστών που είναι εγγεγραμμένοι στην εφαρμογή *forDiabetes*.

Επιπλέον, με την οπτικοποίηση των στατιστικών δειγμάτων των χρηστών που χρησιμοποιούν με συστηματικό τρόπο τους διαφορετικούς τύπους μετρήσεων της εφαρμογής ξεχωριστά αλλά και σε συνδυασμό για ένα ικανό συνεχόμενο χρονικό διάστημα και όχι περιστασιακά, αποδίδεται με εύληπτο τρόπο η ανάγκη διαμόρφωσης κινήτρων στους χρήστες ώστε να χρησιμοποιούν πιο συστηματικά την εφαρμογή, εντάσσοντάς την στην καθημερινότητά τους.

Με την εμπειρία της κατανόησης των παραπάνω, διαμορφώνεται δυο (2) προτάσεις επέκτασης της εφαρμογής. Η πρώτη δίνει τη δυνατότητα οπτικοποίησης σε ένα ενιαίο γράφημα της μεταβολής των τριών βασικών μετρήσεων (γλυκόζη στο αίμα, ημερήσια καύση θερμίδων με την άσκηση, ημερήσια κατανάλωση υδατανθράκων) ώστε ο χρήστης να ερμηνεύει και να κατανοεί εύκολα την αμοιβαία εξάρτησή τους. Η δεύτερη με αναφορά στη βιβλιογραφία αφορά την επέκταση της λειτουργικότητας της εφαρμογής ώστε να καταχωρούνται με συστηματικό τρόπο και μετρήσεις που αφορούν στην ποιότητα του ύπνου του χρήστη. Με την εισαγωγή του ύπνου ως νέο τύπο μέτρησης στην εφαρμογή ο χρήστης θα μπορεί να ελέγχει αποτελεσματικότερα την τιμή της γλυκόζης τους, αφού θα λαμβάνει υπόψη τον παράγοντα ύπνο που μέχρι τώρα μπορεί να αγνοούσε.

6.2 Σημασία της πτυχιακής εργασίας

Η παρούσα πτυχιακή εργασία αποτελεί σημαντικό σταθμό των προπτυχιακών σπουδών μου. Ξεκινώντας ως πληροφορικός με γνώσεις στον προγραμματισμό σε R που πήρα σε πρώτη έκθεση από τα μαθήματα «Εισαγωγή στην Ανάλυση Δεδομένων» και «Οργάνωση Δεδομένων και Εξόρυξη Πληροφορίας», χρειάστηκε να κτίσω πάνω σε αυτές και να τις βελτιώσω σημαντικά σε θέματα τεχνολογίας που έπρεπε να χρησιμοποιήσω στην πτυχιακή εργασία. Συγκεκριμένα την επεξεργασία δεδομένων με τα πακέτα *dplyr* και *tidyr*, την οπτικοποίηση των αποτελεσμάτων με το πακέτο *ggplot2* και τη δημιουργία δυναμικών γραφημάτων με τα πακέτα *flexdashboard* και *shiny*.

Επιπλέον, έμαθα πολλά και ωφελήθηκα αποκτώντας γνώσεις που έχουν να κάνουν με καθαρά ιατρικά ζητήματα. Από αυτή την άποψη η αλληλεπίδραση με την κυρία Ζιώγα Χριστίνα ήταν καθοριστική.

Τελευταία αλλά εξίσου σημαντική ήταν και η αλληλεπίδραση με την εταιρεία Tessera Multimedia. Κατά την εκπόνηση της πτυχιακής εργασίας και με την ολοκλήρωσή της, διεξήχθησαν δύο (2) παρουσιάσεις/συναντήσεις εργασίας, κάθε μία διάρκειας μίας ώρας, περίπου. Τα αποτελέσματα και οι προτάσεις για τη βελτίωση/επέκταση της εφαρμογής *forDiabetes* ετέθησαν υπόψη της ομάδας ανάπτυξης της εφαρμογής και στη διεύθυνση της εταιρείας Tessera Multimedia και συζητήθηκαν με τον πλέον εποικοδομητικό τρόπο. Η αντίδραση της ομάδας ανάπτυξης της εφαρμογής *forDiabetes* και της διεύθυνσης της εταιρείας στις δύο προτάσεις που αναφέρονται στο Κεφάλαιο 5 υπήρξε καθόλα θετική.

Κεφάλαιο 7ο: ΒΙΒΛΙΟΓΡΑΦΙΑ

- [1] eHealth, W., 2021. *Mhealth: New Horizons For Health Through Mobile Technologies: Second Global Survey On Ehealth*. [online] [Apps.who.int](https://apps.who.int). Available at: <<https://apps.who.int/iris/handle/10665/44607>> [Accessed 14 January 2021].
- [2] Dervos, D. and Coleman, A. (2006). A Common Sense Approach to Defining Data, Information and Metadata. Proceedings of the Ninth International Society for Knowledge Organization 2006 Conference, Vienna. June 2006, Edited by C. Swertz. Berlin: Ergon
- [3] McCrank, Lawrence. (2002). *Historical Information Science*. Medford, N.J.: Information Today.
- [4] Guy, M., 2021. *Types Of Data Measurement Scales: Nominal, Ordinal, Interval, And Ratio*. [online] My Market Research Methods. Available at: <<https://www.mymarketresearchmethods.com/types-of-data-nominal-ordinal-interval-ratio/>> [Accessed 14 January 2021].
- [5] Aggarwal, C., 2015. *Data Mining*.
- [6] Kaul, K., Tarr, J., Ahmad, S., Kohner, E. and Chibber, R., 2012. Introduction to Diabetes Mellitus. *Advances in Experimental Medicine and Biology*, pp.1-11
- [7] *Diabetes Care*, 2016. 2. Classification and Diagnosis of Diabetes. 40(Supplement 1), pp.S11-S24.
- [8] *Diabetes Care*, 2011. Diagnosis and Classification of Diabetes Mellitus. 35(Supplement_1), pp.S64-S71.
- [9] *Diabetes Care*, 2012. Standards of Medical Care in Diabetes--2013. 36(Supplement_1), pp.S11-S66.
- [10] R-project.org. 2021. *R: The R Project For Statistical Computing*. [online] Available at: <<https://www.r-project.org/>> [Accessed 15 January 2021].
- [11] Tiobe.com. 2021. *Index | TIOBE - The Software Quality Company*. [online] Available at: <<https://www.tiobe.com/tiobe-index/>> [Accessed 15 January 2021].
- [12] Cran.r-project.org. 2021. *CRAN Repository Policy*. [online] Available at: <<https://cran.r-project.org/web/packages/policies.html>> [Accessed 15 January 2021].
- [13] Tidy.tidyverse.org. 2021. *Tidy Messy Data*. [online] Available at: <<https://tidy.tidyverse.org/>> [Accessed 15 January 2021].
- [14] Ggplot2.tidyverse.org. 2021. *Create Elegant Data Visualisations Using The Grammar Of Graphics*. [online] Available at: <<https://ggplot2.tidyverse.org/>> [Accessed 15 January 2021].

- [15] Dplyr.tidyverse.org. 2021. A Grammar Of Data Manipulation. [online] Available at: <<https://dplyr.tidyverse.org/>> [Accessed 15 January 2021].
- [16] v0.4.1, s., 2021. *Scales Package | R Documentation*. [online] Rdocumentation.org. Available at: <<https://www.rdocumentation.org/packages/scales/versions/0.4.1>> [Accessed 15 January 2021].
- [17] Readxl.tidyverse.org. 2021. *Read Excel Files*. [online] Available at: <<https://readxl.tidyverse.org/>> [Accessed 15 January 2021].
- [18] v1.3.1, r., 2021. *Readr Package | R Documentation*. [online] Rdocumentation.org. Available at: <<https://www.rdocumentation.org/packages/readr/versions/1.3.1>> [Accessed 15 January 2021].
- [19] Shiny.rstudio.com. 2021. *Shiny*. [online] Available at: <<https://shiny.rstudio.com/>> [Accessed 15 January 2021].
- [20] Rmarkdown.rstudio.com. 2021. *R Markdown*. [online] Available at: <<https://rmarkdown.rstudio.com/>> [Accessed 15 January 2021].
- [21] DataFlair. 2021. *Pros And Cons Of R Programming Language - Unveil The Essential Aspects! - Dataflair*. [online] Available at: <<https://data-flair.training/blogs/pros-and-cons-of-r-programming-language/>> [Accessed 15 January 2021].
- [22] Rstudio.com. 2021. *Rstudio*. [online] Available at: <<https://rstudio.com/products/rstudio/>> [Accessed 15 January 2021].
- [23] Rstudio.com. 2021. *Rstudio IDE Features*. [online] Available at: <<https://rstudio.com/products/rstudio/features/>> [Accessed 15 January 2021].
- [24] Soltoff, B., 2021. *The Grammar Of Graphics | Computing For The Social Sciences*. [online] Computing for the Social Sciences. Available at: <<https://cfss.uchicago.edu/notes/grammar-of-graphics/>> [Accessed 15 January 2021].
- [25] Kelesidis K., Dervos D., and Sidiropoulos A.. 2020. Quantifying the Difficulty of Academic Course Modules, in PCI '20: 24th Pan-Hellenic Conference on Informatics, November 20–22, 2020, Athens, GR.ACM, New York, NY, USA, 6 pages
- [26] Touma, C. and Pannain, S., 2011. Does lack of sleep cause diabetes?. *Cleveland Clinic Journal of Medicine*, 78, pp.549-558
- [27] Kuhn, E., Brodan, V., Brodanová, M. and Rysánek, K., 1969. Metabolic reflection of sleep deprivation. *Act Nerv Super*, (11), pp.165–174.
- [28] Spiegel, K., Leproult, R. and Van Cauter, E., 1999. mpact of sleep debt on meta-bolic and endocrine function. *Lancet*, (354), pp.1435–1439.
- [29] Gottlieb, D., Punjabi, N. and Newman, A., 2005. Association of sleep time with diabetes mellitus and impaired glucose tolerance. *Arch Intern Med*, (165), pp.863–867.

- [30] Yaggi, H., Araujo, A. and McKinlay, J., 2006. Sleep duration as a risk factor for the development of type 2 diabetes. *Diabetes Care*, (29), pp.657–661.
- [31] *Diabetes Care*, 2015. Sleep duration and risk of type 2 diabetes: a meta-analysis of prospective studies. (38), pp.529-537.
- [32] Cappuccio, F. and Miller, M., 2010. The epidemiology of sleep and cardiovascular risk and disease. *Aetiology to Public Health*, pp.111–140
- [33] Tasali, E., Leproult, R., Ehrmann, D. and Van Cauter, E., 2008. Slow-wave sleep and the risk of type 2 diabetes in humans. *Proc Natl Acad Sci U S A*, (105), pp.1044–1049.
- [34] Aronsohn, R., Whitmore, H., Van Cauter, E. and Tasali, E., 2010. Impact of untreated obstructive sleep apnea on glucose control in type 2 diabetes. *Am J Respir Crit Care Med*, (181), pp.507–513.