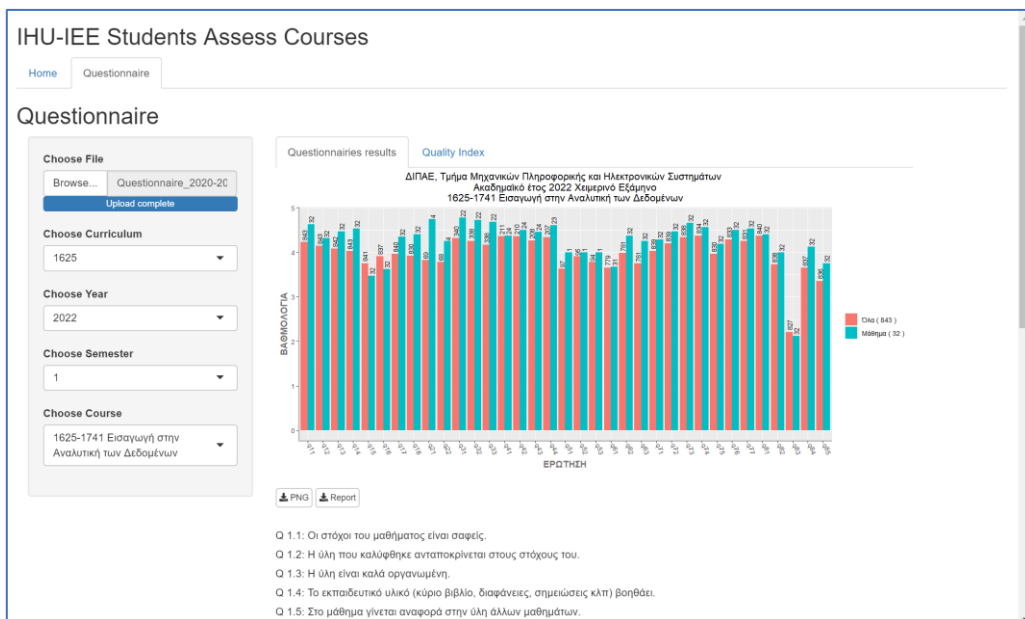


ΣΧΟΛΗ ΜΗΧΑΝΙΚΩΝ
ΤΜΗΜΑ ΜΗΧΑΝΙΚΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ
ΚΑΙ ΗΛΕΚΤΡΟΝΙΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

Υπηρεσία IEE_StAC: Η Τεχνολογία της R στην Πράξη



Φοιτήτρια
ΒΑΣΙΛΙΚΗ ΜΑΡΙΑ ΤΑΝΗ
Αρ. Μητρώου: 123845

Επιβλέπων
ΔΗΜΗΤΡΙΟΣ Α. ΔΕΡΒΟΣ
ΚΑΘΗΓΗΤΗΣ

20 Ιανουαρίου 2023

Τίτλος Π.Ε.: ΥΠΗΡΕΣΙΑ ΙΕΕ_StAC: Η ΤΕΧΝΟΛΟΓΙΑ ΤΗΣ R ΣΤΗΝ ΠΡΑΞΗ

Κωδικός Π.Ε.: 21102

Όνοματεπώνυμο φοιτήτριας: ΒΑΣΙΛΙΚΗ ΜΑΡΙΑ ΤΑΝΗ

Όνοματεπώνυμο εισηγητή ΔΗΜΗΤΡΙΟΣ Α. ΔΕΡΒΟΣ

Ημερομηνία ανάληψης Π.Ε. 25-01-2021

Ημερομηνία περάτωσης Π.Ε. 20-01-2023

Βεβαιώνω ότι είμαι ο 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 διπλωματική εργασία, στο Τμήμα Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων του ΔΙ.ΠΑ.Ε.

Η παρούσα εργασία αποτελεί πνευματική ιδιοκτησία της φοιτήτριας Βασιλικής Μαρίας Τάνη που την εκπόνησε/αν. 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 απαραίτητως και αποδοχή των απόψεων του συγγραφέα, εκ μέρους του Τμήματος.

It is what it is...

Πρόλογος

Από την 4η χιλιετία π.Χ. που ευφυείς Σουμέριοι επινόησαν τη γραφή, και οι άνθρωποι μπορούσαν πλέον να απομνημονεύουν και να επεξεργάζονται πληροφορίες έξω από τον εγκέφαλο, η ποσότητα, η ταχύτητα αλλά και η πολυπλοκότητα διαχείρισης των πληροφοριών εξελίχθηκε σταδιακά και ήπια έως τα τελευταία τριάντα χρόνια, που ο ρυθμός είναι πλέον καταγιστικός. Η επιστήμη της πληροφορικής με εργαλεία τα υπολογιστικά συστήματα επιτελεί μία επανάσταση, μέσα στην οποία ζούμε.

Predictive Analytics, Big Data, Blockchain, Artificial Inteligence, Machine Learning είναι τεχνολογίες που εξελίσσονται στο σήμερα και διαμορφώνουν το αύριο, όχι μόνο της πληροφορικής αλλά και της οικονομίας, της κοινωνίας, της ανθρωπότητας στο σύνολό της.

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 το μάθημα με τίτλο «Οργάνωση Δεδομένων και Εξόρυξη Πληροφορίας». Πριν την παρακολούθηση αυτού του μαθήματος κανένα άλλο μάθημα δεν είχε κινήσει το ενδιαφέρον μου σε τέτοιο βαθμό. Το συγκεκριμένο μάθημα και η συμμετοχή μου στα στάδια επεξεργασίας των πρωτογενών δεδομένων αποτέλεσαν την κινητήρια δύναμη για περαιτέρω διερεύνηση του κλάδου.

Ταυτόχρονα, όντας δραστήρια στην επαγγελματική μου ζωή, κατάφερα να αποκτήσω κάποια επαγγελματική εμπειρία εργαζόμενη σε ομαδικά projects και εργασίες που σχετίζονται με την οργάνωση και διοίκηση. Έτσι διαπίστωσα τη σημασία της βαθιάς κατανόησης της πληροφορίας που μας δίνουν τα δεδομένα, και πώς αυτά μπορούν να αποδειχθούν ένα εξαιρετικό εργαλείο που καθορίζει τη λήψη σημαντικών αποφάσεων. Πρόκληση σε αυτή τη διαδικασία αποτελεί η δυνατότητα κατανόησης της πληροφορίας από επαγγελματίες και επιστήμονες με ελάχιστη ή καθόλου εμπειρία στον προγραμματισμό και τα δεδομένα.

Ευτυχώς, δημιουργικοί επιστήμονες που ασχολούνται με τον προγραμματισμό επινοούν συνεχώς έξυπνες λύσεις. Λύσεις που ενισχύουν τις δυνατότητες μας για αποτελεσματική αξιοποίηση των δεδομένων και της παρουσίας τους με μορφή εύκολα κατανοητή από το κοινό, ανεξάρτητα από το επιστημονικό τους υπόβαθρο.

Σε μία τέτοια σκέψη για βέλτιστη αξιοποίηση των δεδομένων αξιολόγησης των μαθημάτων του Τμήματος, προέκυψε η ιδέα μίας πτυχιακής που θα ασχοληθεί με την τακτοποίηση και την οργάνωση των δεδομένων, καθώς και την οπτικοποίηση της πληροφορίας, με τρόπο που να διευκολύνει την εξαγωγή χρήσιμων συμπερασμάτων.

Η παρούσα πτυχιακή εργασία αποτέλεσε για εμένα προσωπικά λαμπρή ευκαιρία για να γνωρίσω λίγο καλύτερα τον κόσμο της επεξεργασίας, ανάλυσης και οπτικοποίησης των δεδομένων και να εξασκήσω τις δεξιότητές μου, που σχετίζονται με τα εργαλεία και τις τεχνικές επεξεργασίας και οπτικοποίησης δεδομένων, που παρέχει το προγραμματιστικό περιβάλλον της γλώσσας R.

Περίληψη

Η παρούσα πτυχιακή εργασία αποτελείται από δύο βασικούς πυλώνες: την επεξεργασία δεδομένων και την παρουσίαση δεδομένων αξιολόγησης μαθημάτων, όπως αυτά - τα δεδομένα - προέκυψαν από φοιτητές του Τμήματος Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων (ΜΠΗΣ) της Σχολής Μηχανικών του Διεθνούς Πανεπιστημίου της Ελλάδος (ΔΠΠΑΕ) -πρώην Τμήμα Πληροφορικής του Αλεξάνδρειου Τεχνολογικού Εκπαιδευτικού Ιδρύματος Θεσσαλονίκης (ΑΤΕΙΘ).

Ιδιαίτερη έμφαση δίνεται στην οπτικοποίηση της πληροφορίας με φιλικό προς το χρήστη τρόπο, ώστε το εργαλείο και τα αποτελέσματα της χρήσης του, να βοηθήσουν αμφοτέρους καθηγητές και φοιτητές στο πλαίσιο της προσπάθειας που γίνεται για τη βελτίωση της ποιότητας του εκπαιδευτικού έργου του Τμήματος.

Πρωτίστως, τα δεδομένα συλλέχθηκαν από τη βάση δεδομένων του πληροφοριακού συστήματος της ΜΟ.ΔΙ.Π.. Η ΟΜ.Ε.Α. διενεργεί κάθε εξάμηνο τις (ανώνυμες) αξιολογήσεις των μαθημάτων που προσφέρονται στο κάθε ένα από τα προγράμματα σπουδών του Τμήματος. Τα αποτελέσματα των αξιολογήσεων συλλέγονται από το πληροφοριακό σύστημα της ΜΟ.ΔΙ.Π. από όπου και ελήφθησαν. Ακολούθησε η διαδικασία του καθαρισμού, της επεξεργασίας, της ταξινόμησης και της μοντελοποίησης των δεδομένων με στόχο την ερμηνεία τους και την εξαγωγή χρήσιμης πληροφορίας από αυτά.

Ο δεύτερος πυλώνας είναι αυτός της οπτικοποίησης και της παρουσίασης των αποτελεσμάτων της αξιολόγησης. Στην παρούσα εργασία δίνεται ιδιαίτερο βάρος στην οπτικοποίηση της πληροφορίας, και στην παρουσίασή της μέσω μιας φιλικής προς τον χρήστη διαδικτυακής εφαρμογής. Με τη χρήση της τεχνολογίας R και των σχετικών εργαλείων/βιβλιοθηκών όπως η R Shiny, η R Markdown κ.α., αναπτύχθηκε ένα πρωτότυπο λογισμικό, η εφαρμογή/υπηρεσία IEE_StAC. Μέσω αυτής της υπηρεσίας οπτικοποιούνται και παρουσιάζονται στον χρήστη τα αποτελέσματα της αξιολόγησης των μαθημάτων από τους φοιτητές του Τμήματος.

Στα κείμενα που ακολουθούν γίνεται εκτενής αναφορά στην επεξεργασία των δεδομένων, στο λογισμικό που χρησιμοποιήθηκε, και στο παραγόμενο οπτικοποιημένο τελικό αποτέλεσμα. Συχνά, το προϊόν της εργασίας θα αναφέρεται ως «υπηρεσία».

IEE_StAC WEB SERVICE: R TECHNOLOGY IN PRACTICE

VASILIKI MARIA TANI

Abstract

The present final year project thesis report focuses on the processing of academic course module evaluation data. The data originate from questionnaires the students fill in in the middle of each one academic semester at the Information and Electronic Engineering (IEE) department of the International Hellenic University (IHU). The work reported can be seen to comprise of two main part: (a) data collection, preprocessing, and preparation, (b) data visualization. Special emphasis is given to the visualization of information in a user-friendly way that is of assistance both to the academic and the student community in their effort to improve the quality of the educational process.

The data used originate from IHU's Quality Assurance Unit (QU.AS.U.). The IEE department's Internal Evaluation Group (I.E.G.) organize the (anonymous) student evaluation procedure in the middle of each one academic semester. The data are collected by the QU.AS.U. service and they are made available to I.E.G. for further processing. The latter involves a first data cleaning/correction stage, followed by the organization of the data in structures that facilitate subsequent processing. The aim is to first of all interpret the data and its inherent information in order to present the latter in an effective and user-friendly manner.

Particular emphasis is given to the visualization of the information and its availability via a user-friendly web service. A prototype software has been developed and it is presented in this respect, utilizing R/RStudio and a set of R packages/libraries.

The present thesis reports on the data cleaning/correction/preparation stage, the technology and software used, and on the prototype web service developed (IEE_StAC).

Ευχαριστίες

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

Στον επιβλέποντα καθηγητή μου, κ. Δημήτριο Δέρβο, για την πολύτιμη υποστήριξή του, τις παραγωγικές υποδείξεις του, την αστείρευτη ενέργεια του για δημιουργία και συνεχή βελτίωση, και το εξαιρετικό κλίμα συνεργασίας που διαμόρφωσε συμβάλλοντας τα μέγιστα για την εκπόνηση της πτυχιακής μου εργασίας.

Στον υποψήφιο διδάκτορα κ. Κωνσταντίνο Κελεσίδη για την ανεκτίμητη βοήθεια του, και για όλες τις ώρες που αφιέρωσε στην επεξήγηση εννοιών, την διερεύνηση λαθών και την επινότηση λύσεων στον κώδικα R.

Τέλος, στην οικογένεια μου, ιδιαίτερα στην μητέρα μου Ελισσάβη και τον πατέρα μου Κωνσταντίνο, και στους φίλους μου, ιδιαίτερα δε στην Αθανασία, για τη συμπαράσταση και την υπομονή τους.

Περιεχόμενα

Πρόλογος	7
Περίληψη	9
Abstract	11
Ευχαριστίες	13
Περιεχόμενα	15
Κατάλογος Πινάκων	19
Συντομογραφίες	21
Κεφάλαιο 1ο: Εισαγωγή.....	23
1.1 Εσωτερική Αξιολόγηση Τμήματος	23
1.1.1 Ομάδα Εσωτερικής Αξιολόγησης (ΟΜ.Ε.Α.)	23
1.1.2 Μονάδα Διασφάλισης της Ποιότητας (ΜΟ.ΔΙ.Π.)	23
1.2 Αντικείμενο της Υπηρεσίας	24
1.3 Οργάνωση Περιεχομένου.....	24
Κεφάλαιο 2ο: Περιγραφή και Προεπεξεργασία των Δεδομένων	25
2.1 Πηγή Δεδομένων	25
2.2 Ερωτηματολόγια Αξιολόγησης	25
2.2.1 Προγράμματα Σπουδών.....	25
2.2.2 Χρονοδιάγραμμα Αξιολογήσεων	25
2.2.3 Οι Ερωτήσεις του Ερωτηματολογίου	26
2.2.4 Μορφή Απαντήσεων του Ερωτηματολογίου	26
2.3 Δομή Δεδομένων	27
Κεφάλαιο 3ο: Καθαρισμός και Προετοιμασία της Βάσης Δεδομένων	29
3.1 Βήματα Επεξεργασίας των Δεδομένων.....	29
3.1.1 Βήμα 1: Έλεγχος Τίτλων Μαθημάτων	29
3.1.2 Βήμα 2: Διαχωρισμός Μαθημάτων και Επεξεργασία	31
3.1.3 Βήμα 3: Δημιουργία Ενός Νέου Αρχείου .csv για Κάθε Πρόγραμμα Σπουδών	32
3.1.4 Βήμα 4: Μετατροπή Αρχείου .csv σε Αρχείο R data	33
3.2 Τελική Μορφή Δεδομένων.....	33
Κεφάλαιο 4ο: Τεχνολογίες που Χρησιμοποιήθηκαν	35
4.1 RStudio IDE	35
4.2 Γλώσσα R.....	36
4.2.1 Πλεονεκτήματα Γλώσσας R.....	36

4.2.2	Μειονεκτήματα Γλώσσας R	37
4.3	Βιβλιοθήκες R	37
4.4	Βιβλιοθήκη R Shiny	38
4.5	Βιβλιοθήκη R Markdown.....	39
4.5.1	Τα Βασικά Σημεία της R Markdown.....	39
4.5.2	Η Δομή και το Συντακτικό Ενός Αρχείου R Markdown.....	40
4.5.3	Η Γλώσσα Markdown	41
4.5.4	Τα Παραγόμενα Αρχεία	42
4.6	Shinyapps.io	42
4.6.1	Ευχρηστία.....	43
4.6.2	Ασφάλεια.....	43
4.6.3	Ευελιξία.....	43
Κεφάλαιο 5ο:	Οπτικοποίηση Δεδομένων Ερωτηματολογίου	45
5.1	Οπτικοποίηση Δεδομένων.....	45
5.2	Η Γραμματική των Γραφημάτων	46
5.3	Ραβδόγραμμα Αξιολογήσεων.....	47
5.4	Γράφημα Πυκνότητας του Δείκτη Ποιότητας.....	49
5.4.1	Λίγα Λόγια για τον Δείκτη Ποιότητας	51
Κεφάλαιο 6ο:	Η Υπηρεσία IEE_StAC	53
6.1	Με μια Ματιά.....	53
6.2	Λειτουργίες της Υπηρεσίας.....	55
6.2.1	Σύνδεση στην Υπηρεσία	55
6.2.2	Φόρτωση Βάσης Δεδομένων.....	55
6.2.3	Επιλογή Μεταβλητών.....	56
6.2.4	Επιλογή και Εμφάνιση των Γραφημάτων	57
6.2.5	Εκτύπωση των Γραφημάτων και της Αναφοράς του Μαθήματος	57
Κεφάλαιο 7ο:	Προβλήματα που Αντιμετωπίστηκαν και το Μέλλον της Υπηρεσίας	59
7.1	Προβλήματα και Λύσεις.....	59
7.1.1	Τίτλοι και Στοιχεία Μαθημάτων των Δεδομένων	59
7.1.2	Τα Εργαστήρια και η Αξιολόγησή τους.....	59
7.1.3	Διάθεση της Βάσης Δεδομένων	59
7.2	Επόμενα Βήματα	60
Κεφάλαιο 8ο:	Επίλογος	61
Κεφάλαιο 9ο:	Βιβλιογραφία.....	63
ΠΑΡΑΡΤΗΜΑ Α :	Αρχείο Κώδικα R «step1_data_prepare»	65

Κατάλογος Πινάκων

Πίνακας 4.1:.....Βασικά στοιχεία σύνταξης κειμένου της γλώσσας Markdown

Συντομογραφίες

ΔΙΠΑΕ	Διεθνές Πανεπιστήμιο Ελλάδος
Π.Ε.	Πτυχιακή Εργασία
ΜΟ.Δ.Ι.Π.	Μονάδα Διασφάλισης της Ποιότητας
ΟΜ.Ε.Α.	Ομάδα Εσωτερικής Αξιολόγησης
ΒΔ	Βάση Δεδομένων
ΠΣ	Πρόγραμμα Σπουδών

Κεφάλαιο 1ο: Εισαγωγή

Η παρούσα πτυχιακή εργασία έχει ως κύριο στόχο την επεξεργασία και την παρουσίαση δεδομένων αξιολόγησης των μαθημάτων, από φοιτητές του Τμήματος Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων (ΜΠΗΣ) της Σχολής Μηχανικών του Διεθνούς Πανεπιστημίου της Ελλάδος (ΔΠΠΑΕ) -πρώην Τμήμα Πληροφορικής του Αλεξάνδρειου Τεχνολογικού Εκπαιδευτικού Ιδρύματος Θεσσαλονίκης (ΑΤΕΙΘ). Ακόμη, την οπτικοποίηση της πληροφορίας που προκύπτει στην έξοδο της επεξεργασίας και τη διάθεσή της μέσω μιας φιλικής προς τον χρήστη διαδικτυακής εφαρμογής.

Ο στόχος επιτυγχάνεται με τον σχεδιασμό και την ανάπτυξη ενός πρωτότυπου λογισμικού, κάνοντας χρήση της τεχνολογίας R/RStudio και των σχετικών εργαλείων/βιβλιοθηκών. Μέσω του λογισμικού πραγματοποιείται η οπτικοποίηση της πληροφορίας, αφού πρώτα γίνει ο καθαρισμός, η ταξινόμηση και η τελική επεξεργασία των πρωτογενών δεδομένων, ώστε να δημιουργηθεί η κατάλληλη οργάνωσή τους.

Η εργασία αναφέρεται τόσο στο λογισμικό, την χρήση του, και στο παραγόμενο οπτικοποιημένο αποτέλεσμα, όσο και στην ίδια την τεχνολογία με την οποία αυτό δημιουργήθηκε. Ο τίτλος της (Υπηρεσία IEE_StAC: Η τεχνολογία της R στην πράξη) είναι δηλωτικός του περιεχομένου της. Η διαδικτυακή εφαρμογή/υπηρεσία ονομάζεται IHU-IEE Students Assess Courses (IEE_StAC, εν συντομία).

Στη συνέχεια αναφέρονται οι σχετικές, βασικές έννοιες και γίνεται μία σύντομη περιγραφή του περιεχομένου κάθε ενός από τα κεφάλαια της εργασίας.

1.1 Εσωτερική Αξιολόγηση Τμήματος

Η εργασία σχετίζεται με το έργο της Ομάδας Εσωτερικής Αξιολόγησης (ΟΜ.Ε.Α.) του Τμήματος και κατ' επέκταση με τη Μονάδα Διασφάλισης της Ποιότητας (ΜΟ.ΔΙ.Π.) του πανεπιστημίου. Όλα τα δεδομένα που αξιοποιήθηκαν για την εργασία έχουν δοθεί από τη ΜΟ.ΔΙ.Π. ενώ το τελικό παραδοτέο της εργασίας (η υπηρεσία IEE_StAC) πρόκειται να αξιοποιηθεί από τα μέλη της ΟΜ.Ε.Α του Τμήματος.

1.1.1 Ομάδα Εσωτερικής Αξιολόγησης (ΟΜ.Ε.Α.)

Η ΟΜ.Ε.Α. του Τμήματος ΜΠΗΣ είναι υπεύθυνη για τις διαδικασίες αξιολόγησής του Τμήματος, καθώς και για την σύνταξη της Έκθεσης Εσωτερικής Αξιολόγησης. Επιγραμματικά, μεταξύ άλλων, η ΟΜ.Ε.Α. συντονίζει τη διαδικασία συμπλήρωσης ερωτηματολογίων και απογραφικών δελτίων από φοιτητές του Τμήματος, επεξεργάζεται τα δεδομένα και συντάσσει εκθέσεις αξιολόγησης και πιστοποιήσεις χρήσιμες για το Τμήμα και τη Μονάδα Διασφάλισης της Ποιότητας (ΜΟ.ΔΙ.Π.) του Πανεπιστημίου. Με την εφαρμογή που δημιουργήθηκε, η ΟΜ.Ε.Α. αλλά και το ακαδημαϊκό προσωπικό θα μπορούν να κατανοούν τα δεδομένα αξιολόγησης καθώς και να παράγουν διαγράμματα με αυτοματοποιημένο τρόπο [IS1].

1.1.2 Μονάδα Διασφάλισης της Ποιότητας (ΜΟ.ΔΙ.Π.)

Η ΜΟ.ΔΙ.Π. περιγράφεται ως ένα συμβουλευτικό όργανο για τη διοίκηση του Πανεπιστημίου συγκεντρώνοντας χρήσιμες πληροφορίες για το Πανεπιστήμιο και το πως το ίδιο μπορεί να παρέχει υψηλού επιπέδου ποιότητας ακαδημαϊκές εμπειρίες σε φοιτητές και άμεσα εμπλεκόμενους. Η ΜΟ.ΔΙ.Π είναι όργανο του ΔΠΠΑΕ ενώ η ΟΜ.Ε.Α. είναι όργανο του Τμήματος ΜΠΗΣ [AP1] [IS2].

1.2 Αντικείμενο της Υπηρεσίας

Η συγκεκριμένη πτυχιακή εργασία, όπως προαναφέρθηκε, αξιοποιεί και παρουσιάζει χρήσιμη πληροφορία για χρήση από το ακαδημαϊκό προσωπικό του Τμήματος ΜΠΗΣ, με χρήση λογισμικού. Η χρήσιμη πληροφορία αφορά μαθήματα που προσφέρονται από τα τρία προγράμματα σπουδών του Τμήματος (ένα προπτυχιακού και δύο μεταπτυχιακού επιπέδου), και είναι αποτέλεσμα επεξεργασίας και ανάλυσης δεδομένων από ερωτηματολόγια αξιολόγησης. Η αξιολόγηση των μαθημάτων πραγματοποιείται από φοιτητές, κάθε ένα ακαδημαϊκό εξάμηνο και για το κάθε ένα μάθημα ξεχωριστά. Πιο αναλυτικά, μέσω της υπηρεσίας ο χρήστης μπορεί να:

- Επιλέξει και συγκρίνει με διαδραστικό τρόπο, αποτελέσματα αξιολογήσεων από κάθε μάθημα που προσφέρθηκε στα τρία Προγράμματα Σπουδών κατά τα ακαδημαϊκά έτη 2020, 2021 και 2022, στο χειμερινό και το εαρινό εξάμηνο .
- Κατανοήσει την δομή των ερωτηματολογίων αξιολόγησης.
- Παράγει γραφήματα που παρουσιάζουν το αποτέλεσμα για κάθε απάντηση σε ερώτηση του ερωτηματολογίου, σε σύγκριση με τον μέσο όρο των απαντήσεων από το σύνολο των μαθημάτων για κάθε ερώτηση.
- Παράγει γραφήματα που παρουσιάζουν έναν δείκτη ποιότητας από κάθε μάθημα, σε κλίμακα σύγκρισης με τα υπόλοιπα μαθήματα που προσφέρθηκαν στο πρόγραμμα σπουδών του εξαμήνου.
- Εκτυπώνει γραφήματα που επιλέγει ο χρήστης ως αρχεία εικόνων (.png) ώστε αυτός να μπορεί να τα αξιοποιήσει κατά βούληση.
- Εκτυπώνει, με αυτοματοποιημένο τρόπο, κείμενα/αναφορές που αφορούν στην αξιολόγηση μαθημάτων ώστε αυτές να αξιοποιηθούν περαιτέρω από την ΟΜ.Ε.Α., ή/και από τους ίδιους τους διδάσκοντες καθηγητές.

1.3 Οργάνωση Περιεχομένου

Συνοπτικά το περιεχόμενο κάθε κεφαλαίου της παρούσας εργασίας:

Στο Κεφάλαιο 2 περιγράφεται αναλυτικά η δομή των πρωτογενών δεδομένων. Ο αναγνώστης μπορεί να κατανοήσει με ποιον τρόπο παράχθηκαν τα δεδομένα, ποιο είναι το περιεχόμενο των ερωτηματολογίων αξιολόγησης, πως ο κάθε φοιτητής αξιολογεί ένα μάθημα και εν κατακλείδι με ποια μορφή παρέχονται τα δεδομένα από το πληροφοριακό σύστημα της ΜΟ.ΔΙ.Π. Αναφέρονται επίσης οι βασικοί ορισμοί όρων που χρησιμοποιούνται.

Στο Κεφάλαιο 3 απαριθμούνται όλα τα βήματα που πρέπει κάποιος να ακολουθήσει ώστε να επεξεργαστεί τα πρωτογενή δεδομένα, με τέτοιο τρόπο ώστε να τα οργανώσει σε μορφή κατάλληλη προς επεξεργασία από την υπηρεσία IEE_StAC.

Στο Κεφάλαιο 4 αναφέρονται οι τεχνολογίες και τα εργαλεία που χρησιμοποιήθηκαν για την επεξεργασία των δεδομένων, καθώς και για την ανάπτυξη και υλοποίηση της υπηρεσίας.

Στο Κεφάλαιο 5 αναλύονται τα αποτελέσματα της οπτικοποίησης των δεδομένων μέσω δύο τύπων γραφημάτων. Επιπλέον ορίζεται η έννοια του δείκτη ποιότητας των μαθημάτων.

Στο Κεφάλαιο 6 περιγράφεται αναλυτικά το περιβάλλον της εφαρμογής IHU-IEE Students Assess Courses και οι λειτουργίες της.

Στο Κεφάλαιο 7 αναφέρονται τα προβλήματα που παρουσιάστηκαν κατά τη διάρκεια εκπόνησης της πτυχιακής εργασίας, καθώς και επόμενα βήματα που αποσκοπούν στην περαιτέρω ανάπτυξη και βελτίωση της εφαρμογής.

Στα Κεφάλαια 8 και 9, ο αναγνώστης βρίσκει τον Επίλογο και τη Βιβλιογραφία της εργασίας.

Κεφάλαιο 2ο: Περιγραφή και Προεπεξεργασία των Δεδομένων

Η Ανάλυση Δεδομένων είναι η διαδικασία κατανόησης, καθαρισμού, επεξεργασίας και μοντελοποίησης δεδομένων, με στόχο τη δυνατότητα αλίευσης χρήσιμης πληροφορίας, παρουσίαση εμπεριστατωμένων λύσεων σε προβλήματα, και την υποστήριξη της διαδικασίας λήψης αποφάσεων [Α1]. Πρωταρχικό βήμα σε αυτή τη ροή ώστε να προχωρήσει κάποιος στην επεξεργασία δεδομένων, είναι η ανάγνωση και κατανόηση των δεδομένων, όπως αυτά παρέχονται από την πηγή.

2.1 Πηγή Δεδομένων

Τα δεδομένα για τη συγκεκριμένη πτυχιακή προήλθαν από τη βάση δεδομένων του πληροφοριακού συστήματος της ΜΟ.ΔΙ.Π.. Η ΟΜ.Ε.Α. οργανώνει τις αξιολογήσεις (ανώνυμες) των μαθημάτων που προσφέρονται, και τα αποτελέσματα συλλέγονται στο εσωτερικό σύστημα της ΜΟ.ΔΙ.Π. όπου και είναι προσβάσιμα στο σύνολό τους από τα μέλη της ΟΜ.Ε.Α., ενώ για το κάθε ένα μάθημα ξεχωριστά μόνον από τον αντίστοιχο διδάσκοντα-καθηγητή.

2.2 Ερωματολογία Αξιολόγησης

Οι αξιολογήσεις στο Τμήμα Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων πραγματοποιούνται με τη μορφή ερωματολογίων τα οποία απαντώνται από τους εγγεγραμμένους και ενεργούς φοιτητές ανώνυμα, για κάθε ένα μάθημα και για το κάθε ένα πρόγραμμα σπουδών.

2.2.1 Προγράμματα Σπουδών

Το κάθε ένα πρόγραμμα σπουδών του Τμήματος έχει μαθήματα που χωρίζονται σε εξάμηνα. Μερικά από τα μαθήματα περιλαμβάνουν πρακτικό και εργαστηριακό μέρος το οποίο αξιολογείται χωριστά το καθένα. Θα πρέπει να αναφερθεί ότι γι'αυτή την εργασία χρησιμοποιήθηκαν αποκλειστικά αποτελέσματα αξιολόγησης μαθημάτων στο σύνολό τους, και ότι μελετήθηκαν και αξιοποιήθηκαν τα δεδομένα από όλα τα διαθέσιμα προγράμματα σπουδών, τα οποία αναφέρονται παρακάτω.

Το Τμήμα Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων προσφέρει τρία προγράμματα σπουδών, ένα προπτυχιακό (ΠΠΣ) και δύο μεταπτυχιακά (ΠΜΣ).

Οι σπουδές στο Προπτυχιακό Τμήμα Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων (ΠΠΣ) διαρκούν πέντε έτη και συμπεριλαμβάνουν 85 μαθήματα.

Το ΠΜΣ «Ευφυείς Τεχνολογίες Διαδικτύου» αποτελείται από 10 μαθήματα και διαρκεί δύο (2) ακαδημαϊκά έτη.

Το ΠΜΣ «Εφαρμοσμένα Ηλεκτρονικά Συστήματα» αποτελείται από 9 μαθήματα, και διαρκεί δύο (2) ακαδημαϊκά έτη.

2.2.2 Χρονοδιάγραμμα Αξιολογήσεων

Όλα τα μαθήματα που προσφέρονται στο κάθε ένα πρόγραμμα σπουδών διακρίνονται σε μαθήματα του Χειμερινού και σε μαθήματα του Εαρινού ακαδημαϊκού εξαμήνου. Το Χειμερινό εξάμηνο έχει έναρξη τον Σεπτέμβριο και λήξη τον Ιανουάριο, ενώ το Εαρινό αρχίζει τον Φεβρουάριο και ολοκληρώνεται τον Ιούλιο του κάθε ενός ημερολογιακού έτους. Οι αξιολογήσεις των μαθημάτων είναι ανώνυμες και γίνονται στο μέσον (περίπου) κάθε εξαμήνου και αφορούν στο κάθε ένα από τα μαθήματα που προσφέρονται το συγκεκριμένο ακαδημαϊκό εξάμηνο. Ο εκάστοτε διδάσκων έχει ευθύνη να προωθήσει την αξιολόγηση του μαθήματος του στους φοιτητές που παρακολουθούν το μάθημα σε συγκεκριμένο

χρόνο. Η αξιολόγηση γίνεται διαδικτυακά μέσω λογισμικού της ΜΟ.ΔΙ.Π. με τη χρήση κωδικών οι οποίοι δίνονται στους φοιτητές από τον εκάστοτε διδάσκοντα [IS2].

2.2.3 Οι Ερωτήσεις του Ερωτηματολογίου

Τα ερωτηματολόγια των αξιολογήσεων κάθε Προγράμματος Σπουδών ακολουθούν την ίδια δομή όσον αφορά στις ερωτήσεις που εμπεριέχονται. Οι ερωτήσεις είναι στο σύνολο τους 35 και χωρίζονται σε οκτώ κατηγορίες:

- **Μάθημα:** Αφορά τους στόχους και το περιεχόμενο του μαθήματος (οκτώ ερωτήσεις).
- **Φροντιστηριακά μαθήματα:** Σε κάποια από τα μαθήματα που προσφέρονται δημιουργούνται φροντιστηριακά τμήματα που παρέχουν υποστήριξη σε φοιτητές. Αυτή η κατηγορία απαντήσεων απαντάται μόνο αν το μάθημα προσφέρει φροντιστηριακά μαθήματα (δύο ερωτήσεις).
- **Ασκήσεις πράξης:** Κάποια μαθήματα εμπεριέχουν επιπλέον ασκήσεις πράξης, ανεξάρτητα του εργαστηρίου, αν υπάρχει. Αυτή η κατηγορία απαντήσεων απαντάται μόνο αν το μάθημα προσφέρει φροντιστηριακά μαθήματα (τρεις ερωτήσεις).
- **Εργασίες:** Ερωτήσεις που αφορούν τυχόν εργασίες, ατομικές ή ομαδικές, αν δόθηκαν από τον διδάσκοντα (τέσσερις ερωτήσεις).
- **Εξέταση του μαθήματος:** Οι ερωτήσεις αυτές αφορούν το περιεχόμενο της εξέτασης του μαθήματος. Για να απαντηθούν απαιτείται η αξιολόγηση να γίνει μετά την εξέταση του μαθήματος (τρεις ερωτήσεις).
- **Εκπαιδευτικές υποδομές:** Εδώ αξιολογούνται οι υποδομές και οι τεχνολογίες που χρησιμοποιήθηκαν για το μάθημα (τρεις ερωτήσεις).
- **Ο διδάσκων:** Ο φοιτητής αξιολογεί τον καθηγητή του μαθήματος για τον τρόπο διδασκαλίας, τις μεθόδους του, την συνέπεια του, κ.α. (επτά ερωτήσεις).
- **Εγώ ο φοιτητής:** Σε αυτήν την ομάδα ερωτήσεων ο φοιτητής κάνει την αυτοκριτική του, σε σχέση με τη συμμετοχή του στο μάθημα και στην μελέτη του γι' αυτό (πέντε ερωτήσεις).

2.2.4 Μορφή Απαντήσεων του Ερωτηματολογίου

Ο φοιτητής μπορεί να αξιολογήσει το κάθε μάθημα στο οποίο συμμετείχε, βαθμολογώντας την κάθε ερώτηση σε κλίμακα 1-5, «1» είναι η χαμηλότερη βαθμολογία, «Διαφωνώ απόλυτα» και «5» η υψηλότερη, «Συμφωνώ Απόλυτα». Για παράδειγμα, στην ερώτηση «q11: Οι στόχοι του μαθήματος είναι σαφείς.» το «5» αντιστοιχεί στο «Συμφωνώ απόλυτα» όπως φαίνεται και στην Εικόνα 2.1. Εφόσον ο φοιτητής απαντήσει στις ερωτήσεις του ερωτηματολογίου, καταθέτει την αξιολόγηση του. Τότε, τα δεδομένα της αξιολόγησης καταχωρούνται στη βάση δεδομένων της ΜΟ.ΔΙ.Π. [IIS].

Μάθημα	
Οι στόχοι του μαθήματος είναι σαφείς.	5 - Συμφωνώ απόλυτα
Η ύλη που καλύφθηκε ανταποκρίνεται στους στόχους του μαθήματος.	4 - Συμφωνώ
Η ύλη είναι καλά οργανωμένη.	3 - Ούτε συμφωνώ ούτε διαφωνώ
Το εκπαιδευτικό υλικό (κύριο βιβλίο, διαφάνειες, σημειώσεις κλη) βοηθάει στην κατανόηση του μαθήματος.	2 - Διαφωνώ
Στο μάθημα γίνεται αναφορά στην ύλη άλλων μαθημάτων.	1 - Διαφωνώ απόλυτα
Στο μάθημα γίνεται σύνδεση γνώσεων με άλλα μαθήματα.	5 - Συμφωνώ απόλυτα
Το επίπεδο δυσκολίας του μαθήματος είναι ανάλογο του έτους του.	4 - Συμφωνώ
Ο αριθμός των Διδακτικών Μονάδων του μαθήματος είναι ανάλογος με το φόρτο εργασίας.	3 - Ούτε συμφωνώ ούτε διαφωνώ

Εικόνα 2.1: Τμήμα του ερωτηματολογίου αξιολόγησης από τη σελίδα της ΜΟ.ΔΙ.Π.

2.3 Δομή Δεδομένων

Σε κάθε εξάμηνο ένας νέος κύκλος αξιολογήσεων ανοίγει, και ολοκληρώνεται στο τέλος του εξαμήνου. Όπως αναφέρθηκε, τα δεδομένα είναι αποθηκευμένα στη βάση δεδομένων της ΜΟ.ΔΙ.Π. και μόνο εξουσιοδοτημένοι χρήστες έχουν πρόσβαση σε αυτά. Μετά το πέρας της περιόδου αξιολογήσεων τα πρωτογενή αυτά δεδομένα μεταβιβάζονται στην ΟΜ.Ε.Α ώστε να αξιοποιηθούν.

Τα δεδομένα μπορούν να αποθηκευτούν τοπικά σε διάφορους τύπους αρχείων όπως .csv, .xlsx κ.α.. Για τις ανάγκες της πτυχιακής τα δεδομένα οργανώθηκαν σε αρχεία .csv. Κάθε αποθηκευμένο αρχείο εμπεριέχει όλα τα δεδομένα που συλλέχθηκαν από τις αξιολογήσεις των μαθημάτων για το κάθε ένα ΠΣ, ξεχωριστά. Στην Εικόνα 2.2 φαίνεται η δομή των δεδομένων στην πρωτογενή τους μορφή, όπως ανακτώνται από το πληροφοριακό σύστημα της ΜΟ.ΔΙ.Π..

Εισαγωγή στην Αναλυτική των Δεδομένων (1625-1741, θεωρία) προπτυχιακό									
Επίπεδο: προπτυχιακό									
Διδάσκων: Δ. Δέρβος									
Χρήστης: dad									
q11. Οι στόχοι του μαθήματος είναι σαφείς.	5	5	5	4	4	5	4	5	
q12. Η ύλη που καλύφθηκε ανταποκρίνεται σ	5	5	5	4	4	5	3	5	
q13. Η ύλη είναι καλά οργανωμένη.	5	5	5	4	4	5	4	5	
q14. Το εκπαιδευτικό υλικό (κύριο βιβλίο, δια	5	5	5	4	5	5	4	5	
q15. Στο μάθημα γίνεται αναφορά στην ύλη	2	5	5	4	3	5	4	5	
q16. Στο μάθημα γίνεται σύνδεση γνώσεων μ	4	5	5	4	3	5	3	5	
q17. Το επίπεδο δυσκολίας του μαθήματος ε	5	5	5	5	4	5	5	5	
q18. Ο αριθμός των Διδακτικών Μονάδων το	5	5	5	5	4	5	4	5	
q21. Τα φροντιστηριακά μαθήματα βοηθούν στην κατανόηση του μαθήματος.									
q22. Τα φροντιστηριακά μαθήματα ενισχύουν την επίδοσή μου.									
q31. Οι Ασκήσεις-Πράξεις βοηθούν στην κατ	5	5					4	5	

Εικόνα 2.2: Τμήμα του αρχείου δεδομένων .csv μαθήματος του ΠΠΣ

Στο κάθε ένα μάθημα αναλογούν συγκεκριμένες σειρές στη βάση. Η πρώτη σειρά για τον τίτλο του μαθήματος, η δεύτερη σειρά για το επίπεδο σπουδών, η τρίτη σειρά για το όνομα του διδάσκοντα, η τέταρτη σειρά για το όνομα χρήστη και 35 σειρές για τις ερωτήσεις/απαντήσεις. Στις στήλες του αρχείου αναγράφονται οι απαντήσεις του κάθε ενός φοιτητή. Σε κάθε στήλη (εξαιρουμένης της πρώτης- στην οποία αναγράφονται οι ερωτήσεις του ερωτηματολογίου) αντιστοιχίζονται οι απαντήσεις ενός μοναδικού φοιτητή. Στην Εικόνα 2.2 μπορεί κάποιος να αναγνωρίσει ότι για το μάθημα «Εισαγωγή στην Αναλυτική των Δεδομένων» μάλλον δεν είχαν διεξαχθεί (έως την χρονική στιγμή που συμπληρώθηκαν τα ερωτηματολόγια) φροντιστηριακού τύπου διαλέξεις (“ασκήσεις-πράξεις”, ερώτηση q31), καθώς μερικές από τις απαντήσεις είναι κενές. Βέβαια, στη συγκεκριμένη περίπτωση, εντοπίζεται και μία ασάφεια στη διατύπωση αυτής καθεαυτής της ερώτησης στο ερωτηματολόγιο: οι φοιτητές που βαθμολόγησαν την ερώτηση q31 θεώρησαν “ασκήσεις-πράξεις” τις ασκήσεις που τους ζητήθηκε να εκπονήσουν στο σπίτι (homework).

Τα πρωτογενή δεδομένα δεν συνιστούν εύληπτη/χρήσιμη πληροφορία για τον χρήστη. Μετά από μία πρώτη ερμηνεία/κατανόησή τους, μπορεί κάποιος να περάσει στο επόμενο στάδιο το οποίο αφορά στον καθαρισμό και την επεξεργασία των πρωτογενών δεδομένων. Στο επόμενο κεφάλαιο παρουσιάζεται η επεξεργασία των δεδομένων που έγινε στο πλαίσιο της παρούσας πτυχιακής εργασίας.

Κεφάλαιο 3ο: Καθαρισμός και Προετοιμασία της Βάσης Δεδομένων

Στόχος της Επιστήμης των Δεδομένων είναι να μετατρέπει τα δεδομένα σε πληροφορία και την πληροφορία σε γνώση [B1]. Αυτός ο στόχος θα μπορούσε κάποιος να πει ότι είναι αναπόσπαστο κομμάτι σε κάθε διαδικασία που εμπεριέχει ανάλυση δεδομένων. Τα «ποιοτικά δεδομένα», αναμφισβήτητα προσφέρουν πληροφορία και σε μετέπειτα στάδιο γνώση [A3], αρκεί να αξιοποιηθούν σωστά. Τα δεδομένα, για να μπορέσουν να χαρακτηριστούν ως ποιοτικά, οφείλουν να είναι ακριβή, πλήρη, συνεπή και αξιόπιστα [A4]. Αυτά τα στοιχεία πρέπει να χαρακτηρίζουν και τα δεδομένα της συγκεκριμένης εργασίας.

Τα δεδομένα της εργασίας χαρακτηρίζονται αξιόπιστα καθώς λήφθηκαν από αξιόπιστη πηγή, αλλά και συνεπή καθώς λαμβάνονται τακτικά κάθε εξάμηνο. Για να περιγραφούν όμως ως ακριβή και πλήρη πρέπει να ελεγχθούν και να υποστούν επεξεργασία ώστε να πάρουν την μορφή που χρειάζεται και να μπορέσουν να αξιοποιηθούν από ένα λογισμικό.

Σε αυτό το κεφάλαιο αναλύονται τα βήματα που ακολουθήθηκαν για τον έλεγχο, τον καθαρισμό και την μορφοποίηση των δεδομένων και παρουσιάζεται ο σχηματισμός της μορφής τους για χρήση από το λογισμικό.

3.1 Βήματα Επεξεργασίας των Δεδομένων

Τα πρωτογενή δεδομένα είναι αποθηκευμένα σε αρχεία όπου το καθένα περιέχει όλα τα μαθήματα ανά πρόγραμμα σπουδών. Έτσι, υπάρχουν τρία ξεχωριστά αρχεία, ένα για κάθε ΠΣ. Κατά τη διάρκεια της συγγραφής της εργασίας εξετάστηκαν πολλά σενάρια για το ποιος θα ήταν ο βέλτιστος τρόπος μορφοποίησης των δεδομένων και η βέλτιστη τελική τους μορφή.

Κάποια από τα ερωτηματικά ήταν αν θα πρέπει η βάση δεδομένων να αποτελείται από ένα ενιαίο αρχείο ή από ξεχωριστά αρχεία, αν η επεξεργασία δεδομένων θα πρέπει να γίνεται από το ίδιο λογισμικό ή αν θα πρέπει να υφίστανται επεξεργασία εκτός του λογισμικού ώστε αυτό να τα λαμβάνει έτοιμα, κ.α.. Εν τέλει, ακολουθήθηκε ο παρακάτω δρόμος ο οποίος καλύπτει κάθε ανάγκη στην παρούσα περίοδο.

3.1.1 Βήμα 1: Έλεγχος Τίτλων Μαθημάτων

Οι ονομασίες όλων των μαθημάτων και στα τρία προγραμμάτων σπουδών έχουν καθοριστεί κατά τη δημιουργία του ΠΣ ή σε αλλαγές που γίνονται ανά τακτά χρονικά διαστήματα από το Τμήμα της σχολής.

Κάθε καθηγητής, πριν την έναρξη της αξιολόγησης που γίνεται ανά εξάμηνο, είναι υπεύθυνος να δημιουργήσει την αξιολόγηση του μαθήματος του στο πληροφοριακό σύστημα της ΜΟ.ΔΙ.Π.. Οφείλει να περάσει όλα τα στοιχεία που απαιτούνται, όπως το εξάμηνο που διδάσκεται το μάθημα, το ΠΣ στο οποίο ανήκει το μάθημα, το όνομα του καθηγητή, τον κωδικό μαθήματος, κλπ., με μία συγκεκριμένη κωδικοποίηση που έχει ορίσει η ΟΜ.Ε.Α. του Τμήματος.

Σε αυτό το στάδιο είναι δύσκολο να αποφευχθεί το ανθρώπινο λάθος ακόμη και αν οι οδηγίες της ΟΜ.Ε.Α. είναι αναλυτικές. Λάθη που φαίνεται να συμβαίνουν συχνά είναι η χρήση κεφαλαίων γραμμάτων αντί μικρών, λάθη απροσεξίας στον τίτλο του μαθήματος, κωδικοί μαθημάτων που δεν ανταποκρίνονται στην πραγματικότητα, κ.α.. Αν κάποιο από τα παραπάνω λάθη συμβεί τότε τα δεδομένα μας δεν είναι ακριβή και όπως προαναφέρθηκε η ακρίβεια είναι ένα από τα απαραίτητα χαρακτηριστικά των ποιοτικών δεδομένων.

Στην Εικόνα 3.1 δίνεται παράδειγμα από το πως φαίνεται ένα μάθημα το οποίο περάστηκε από καθηγητή στο σύστημα με μη ορθά στοιχεία.

3.1.2 Βήμα 2: Διαχωρισμός Μαθημάτων και Επεξεργασία

Μετά το πρώτο βήμα, το περιεχόμενο των δεδομένων έχει ελεγχθεί για την ορθότητα του και οι ονομασίες και τίτλοι μαθημάτων είναι ολοκληρωμένοι. Στο Βήμα 2 αρχίζει η μετατροπή της δομής της βάσης δεδομένων.

Πλέον, η διαδικασία επεξεργασίας γίνεται με τη βοήθεια της γλώσσας R και των βιβλιοθηκών που προσφέρει στο περιβάλλον του Rstudio. Τα εργαλεία αυτά περιγράφονται και αναλύονται στο επόμενο κεφάλαιο (Κεφάλαιο 4). Γι' αυτό το βήμα έχει δημιουργηθεί ένα script R «step1_data_prepare.R», το οποίο παρουσιάζεται στο Παράρτημα Α, με το οποίο γίνεται η πρώτη φάση της επεξεργασίας.

Πριν από οποιαδήποτε κίνηση, ο χρήστης ορίζει το 'working directory' του, στον φάκελο που βρίσκεται το αρχείο, κάνοντας χρήση της εντολής `setwd('')`. Ύστερα, τα τρία αρχεία θα πρέπει να περάσουν ένα προς ένα από επεξεργασία με την εντολή `Data <- read_xlsx("/22-23_Winter_PPS.xlsx" , col_names =FALSE)`.

Σε αυτό το σημείο στόχος είναι ο διαχωρισμός του ενός αρχείου για όλα τα μαθήματα ενός ΠΣ σε πολλά μοναδικά αρχεία, ένα για κάθε μάθημα. Με αυτή την τακτική η επεξεργασία είναι πιο εύκολη και κατανοητή, καθώς κάθε μάθημα αποτελεί μια διαφορετική οντότητα την οποία μπορεί κάποιος να αναπλάσει και να αξιοποιήσει.

Για να τμηματοποιηθεί το αρχείο πρέπει πρώτα να αναγνωριστεί το μοτίβο (pattern) επανάληψης στο αρχείο που εμπεριέχει όλα τα μαθήματα. Δηλαδή, να γίνει κατανοητό σε ποια σημεία στη βάση γίνεται η αλλαγή μετάβασης από το ένα μάθημα στο επόμενο.

Στο Κεφάλαιο 2.3 εξηγήθηκε η δομή του αρχείου και ότι για κάθε μάθημα αναλογούν συγκεκριμένες σειρές στη βάση. Με μία ανάγνωση του αρχείου γίνεται φανερό ότι κάποιες από τις σειρές της πρώτης στήλης όπου βρίσκονται οι ερωτήσεις, επαναλαμβάνονται. Η τελευταία ερώτηση για κάθε μάθημα είναι η q85, με αυτή την πληροφορία δημιουργείται ένα μοτίβο και με την εντολή `index <- which(grepl("q85", Data$...1))` χωρίζεται η πληροφορία σε τμήματα.

Στη συνέχεια, με σειρά εντολών, κάθε τμήμα ελέγχεται και διαμορφώνεται οπότε με διάφορες παρεμβάσεις παίρνει την τελική μορφή. Φτάνοντας στο τελευταίο τμήμα κώδικα αυτού του βήματος, Εικόνα 3.3, αναγνωρίζεται αν το τμήμα της βάσης είναι «θεωρία» ή «εργαστήριο» και αποθηκεύεται τοπικά με την ανάλογη ονομασία.

```
if(testdata[1,3]=="θεωρία") {  
  path <- paste0(mydir,"/",title,"--theory.csv")  
}  
if(testdata[1,3]=="εργαστήριο") {  
  path <- paste0(mydir,"/",title,"--lab.csv")  
}  
write.csv(testdata, path, row.names = FALSE)
```

Εικόνα 3.3: Τμήμα κώδικα για την αποθήκευση μαθημάτων ανάλογα με το αν είναι 'θεωρία' ή 'εργαστήριο' σε γλώσσα R.

Στο παραπάνω τμήμα κώδικα φαίνεται ότι τα νέα αρχεία είναι τύπου .csv. Έτσι, έχουν δημιουργηθεί ξεχωριστά αρχεία για κάθε μάθημα, θεωρητικό ή εργαστηριακό, τα οποία έχουν τοποθετηθεί σε έναν φάκελο. Η ίδια διαδικασία θα πρέπει να επαναληφθεί και για τα άλλα δύο Προγράμματα Σπουδών.

3.1.3 Βήμα 3: Δημιουργία Ενός Νέου Αρχείου .csv για Κάθε Πρόγραμμα Σπουδών

Γι' αυτό το βήμα δημιουργήθηκε ένα δεύτερο αρχείο κώδικα (script), το «step2_data_prepare», το οποίο παρουσιάζεται στο Παράρτημα Β, το οποίο θα χρησιμοποιηθεί στο επόμενο στάδιο επεξεργασίας των δεδομένων. Αξίζει να αναφερθεί πως ο λόγος που δημιουργήθηκαν δύο διαφορετικά αρχεία κώδικα είναι για την ευκολία στην κατανόηση. Ο κώδικας στο σύνολο του θα μπορούσε να τοποθετηθεί σε ένα αρχείο το οποίο βέβαια θα είχε μεγαλύτερη έκταση και θα ήταν ίσως πιο δύσκολο να διαβαστεί και να κατανοηθεί από τον άνθρωπο.

Στο προηγούμενο βήμα δημιουργήθηκαν ξεχωριστά αρχεία για κάθε μάθημα και αυτά αποθηκεύτηκαν σε ξεχωριστούς φακέλους ανάλογα με το πρόγραμμα σπουδών στο οποίο ανήκουν. Με το δεύτερο τμήμα κώδικα, αρχικά, εξετάζεται σε ποιο ΠΣ αναλογεί ο κάθε φάκελος αλλά και τα μαθήματα που βρίσκονται μέσα σε αυτόν. Στη συνέχεια, τα μαθήματα χαρακτηρίζονται με συγκεκριμένη κωδικοποίηση, ανάλογα με το ΠΣ στο οποίο ανήκουν, και οι κωδικοί αποθηκεύονται στη μνήμη με τη μορφή μεταβλητών.

Ακολούθως, καταγράφονται οι διευθύνσεις των μαθημάτων στη μνήμη με τη μορφή λίστας -ξεχωριστή λίστα για θεωρίες και ξεχωριστή για τα εργαστήρια- με την εντολή `Theories = list.files(path=mydir, pattern="*--theory.csv", full.names=TRUE)` για τις θεωρίες, και ανάλογη εντολή εκτελείται για τα εργαστήρια. Με την ευκαιρία αναφέρεται ότι τα εργαστήρια δε χρησιμοποιούνται στην τελική οπτικοποίηση των δεδομένων της εργασίας αλλά η επεξεργασία τους γίνεται όπως και για τις θεωρίες.

Πλέον, στη μνήμη υπάρχουν δύο λίστες με αποθηκευμένες διευθύνσεις αρχείων, μία για τις θεωρίες και μία για τα εργαστήρια. Για κάθε λίστα δημιουργήθηκε ένα αντίστοιχο τμήμα κώδικα που επιτυγχάνει τα παρακάτω για κάθε μάθημα:

- Διαβάζει το αρχείο του μαθήματος
- Αναγνωρίζει που βρίσκονται οι ερωτήσεις και ο τίτλος του μαθήματος στα δεδομένα και προσαρμόζει τον τίτλο τους.
- Μετατρέπει όλα τα δεδομένα που αφορούν τις απαντήσεις των ερωτήσεων (q) σε «numeric», αριθμητικά πεδία.
- Αλλάζει τη δομή του αρχείου μετατρέποντας τις λίστες σε σειρές και τις σειρές σε λίστες.
- Βρίσκει τον Μ.Ο. απαντήσεων για όλες τις ερωτήσεις του μαθήματος
- Υπολογίζει το πλήθος απαντήσεων στο ερωτηματολόγιο για κάθε μάθημα.

Εφόσον οι παραπάνω πληροφορίες αποθηκευτούν στη μνήμη του προγράμματος, το επόμενο στάδιο είναι να βρεθούν οι Μ.Ο. και το πλήθος απαντήσεων για το σύνολο των μαθημάτων, ξεχωριστά για θεωρίες και για εργαστήρια.

Στη μνήμη υπάρχουν πλέον όλα τα στοιχεία που θεωρούνται απαραίτητα για την οπτικοποίηση των δεδομένων που αφορούν τους Μ.Ο. των απαντήσεων ερωτηματολογίων και τους δείκτες ποιότητας μαθημάτων στη συγκεκριμένη εργασία. Βέβαια, εκτός των αριθμητικών δεδομένων, είναι απαραίτητο να δοθούν σωστές ονομασίες στις επικεφαλίδες στηλών της βάσης, που θα χρησιμοποιηθούν για την αναπαράσταση της πληροφορίας στην εφαρμογή, Εικόνα 3.4.

```
colnames(df)<-c("Question", "Title", "Mean", "C_counter", "AllMean", "A_Counter")
```

Εικόνα 3.4: Παραδείγματα ονοματοδοσίας κάποιων από τις στήλες της ΒΔ σε γλώσσα R.

Τέλος, τα δεδομένα που προκύπτουν μετά από την επεξεργασία και υπάρχουν στη μνήμη του προγράμματος, αποθηκεύονται τοπικά στη μνήμη του υπολογιστή σε ένα αρχείο για κάθε Π.Σ. που περιέχει όλα τα μαθήματα του εξαμήνου.

3.1.4 Βήμα 4: Μετατροπή Αρχείου .csv σε Αρχείο R data

Έως τώρα οι τύποι αρχείων δεδομένων που έχουν χρησιμοποιηθεί για την αποθήκευση των δεδομένων είναι τύπου .csv ή .xlsx. Αυτοί οι τύποι είναι εύχρηστοι για σχετικά μικρές βάσεις δεδομένων [B5]. Καθώς τα εξάμηνα πληθαίνουν, το μέγεθος της βάσης δεδομένων θα αυξάνεται. Αυτό σημαίνει ότι η προσπέλαση της βάσης εξελίσσεται σε μία χρονοβόρα διαδικασία. Στη συγκεκριμένη περίπτωση, η ανάγνωση των δεδομένων από μία εφαρμογή σχεδιασμένη σε γλώσσα R διαρκεί κάποια δευτερόλεπτα, ένας αντικειμενικά μεγάλος χρόνος αναμονής για τον χρήστη της εφαρμογής.

Μετά από δοκιμές διαπιστώθηκε ότι ο τύπος δεδομένων που θα εξυπηρετούσε τις ανάγκες αυτής της ΒΔ και της εφαρμογής είναι ο τύπος που δίνεται από τη γλώσσα R, το αρχείο R data με την προέκταση ονόματος αρχείου .rds. Ένα από τα χαρακτηριστικά του τύπου αρχείου .rds είναι ότι καταλαμβάνει μικρότερο χώρο αποθήκευσης σε σύγκριση με το αντίστοιχο αρχείο σε μορφή .csv ή .xlsx, καθώς υπόκειται σε μεγαλύτερη συμπίεση κατά την αποθήκευση του [IS10].

Στην εργασία, για να αποθηκευτεί η βάση σε μορφή .rds εκτελείται η εντολή *saveRDS(dataall, respath)*, κατά την οποία τα δεδομένα από όλα τα προγράμματα σπουδών για τα εξάμηνα (dataall) αποθηκεύονται τοπικά στην επιθυμητή μορφή, σε ένα ενιαίο αρχείο.

3.2 Τελική Μορφή Δεδομένων

Η βάση δεδομένων στην τελική της μορφή αποτελείται από ένα αρχείο που περιέχει όλα τα επεξεργασμένα στοιχεία.

Title	Category	variable	value	variable2	value2	ProgSpoudon	Year
1625-1101 Μαθηματικά I	θεωρία	Mean	4.039216	C_counter	51	1625	2020
1625-1102 Δομημένος Προγραμματισμός	θεωρία	Mean	4.398148	C_counter	108	1625	2020
1625-1103 Εισαγωγή στην Επιστήμη των Υπολογιστών	θεωρία	Mean	4.277778	C_counter	90	1625	2020
1625-1104 Ηλεκτρονική Φυσική	θεωρία	Mean	3.437500	C_counter	32	1625	2020
1625-1105 Κυκλώματα Συνεχούς Ρεύματος	θεωρία	Mean	4.551402	C_counter	107	1625	2020
1625-1301 Θεωρία Πιθανοτήτων και Στατιστική	θεωρία	Mean	4.122807	C_counter	57	1625	2020
1625-1302 Μαθηματικά III	θεωρία	Mean	4.169811	C_counter	53	1625	2020
1625-1303 Επεξεργασία Σήματος	θεωρία	Mean	3.706897	C_counter	58	1625	2020
1625-1305 Δομές Δεδομένων και Ανάλυση Αλγορίθμων	θεωρία	Mean	3.742857	C_counter	35	1625	2020
1625-1405 Γλώσσες και Τεχνολογίες Ιστού	θεωρία	Mean	4.242857	C_counter	70	1625	2020
1625-1501 Ασύρματες Επικοινωνίες	θεωρία	Mean	4.000000	C_counter	28	1625	2020

Εικόνα 3.5: Κομμάτι της ΒΔ στην τελική της μορφή.

Η βάση δεδομένων, την στιγμή που γράφεται η εργασία, αποτελείται από 13 πεδία και 15.376 σειρές δεδομένων.

Παρακάτω αναλύονται τα πεδία της βάσης.

- **QN:** Δίνεται ο κωδικός ερώτησης όπως q11, q12, q32.
- **Question:** Παρουσιάζεται η ερώτηση ολογράφως, όπως «Οι στόχοι του μαθήματος είναι σαφείς»
- **Title:** Δίνεται ο τίτλος του μαθήματος ο οποίος αποτελείται από τον κωδικό και το όνομα του μαθήματος, π.χ. «1625-1944 Διαχείριση Συστήματος και Υπηρεσιών DBMS»

- **Category:** Σε αυτό το πεδίο καθορίζεται αν το μάθημα είναι θεωρία ή εργαστήριο.
- **Variable:** Αυτό το πεδίο αφορά τον Μ.Ο. σε μία απάντηση ερώτησης για ένα μάθημα και αντίστοιχα για όλα τα μαθήματα στην ίδια ερώτηση. Η μεταβλητή μπορεί να πάρει την τιμή «Mean» ή «AllMean». Το πρώτο αφορά ένα μάθημα ενώ το δεύτερο όλα τα μαθήματα του ίδιου προγράμματος σπουδών. Αυτή η μεταβλητή είναι άμεσα συνδεδεμένη με την επόμενη μεταβλητή «Value».
- **Value:** Εδώ δίνεται η αριθμητική τιμή του Μ.Ο. του προηγούμενου πεδίου «Variable».
- **Variable2:** Αυτό το πεδίο αφορά το πλήθος απαντήσεων ερώτησης για ένα μάθημα και αντίστοιχα για όλα τα μαθήματα στην ίδια ερώτηση. Η μεταβλητή μπορεί να πάρει την τιμή «C_counter» ή «A_Counter». Το πρώτο αφορά ένα μάθημα ενώ το δεύτερο όλα τα μαθήματα του ίδιου προγράμματος σπουδών. Αυτή η μεταβλητή είναι άμεσα συνδεδεμένη με την επόμενη μεταβλητή «Value2».
- **Value2:** Εδώ δίνεται η αριθμητική τιμή του πλήθους του προηγούμενου πεδίου «Variable2».
- **ProgSpoudon:** Σε αυτό το πεδίο αναφέρεται το Πρόγραμμα Σπουδών στο οποίο ανήκει το στοιχείο. Οι τιμές που μπορεί να έχει είναι «1625», «1268» και «1269» για τα τρία προγράμματα σπουδών.
- **Year:** Δίνεται η χρονιά που διδάχθηκε το μάθημα. Οι τιμές μπορεί να είναι 2020, 2021 ή 2022.
- **Semester:** Δίνεται το εξάμηνο που διδάχθηκε το μάθημα. Οι τιμές εδώ είναι και πάλι αριθμητικές 1 ή 2 για το χειμερινό και εαρινό εξάμηνο αντίστοιχα.
- **Id:** Σε αυτή τη στήλη αναφέρεται ο κωδικός του μαθήματος όπως «1269-M103».
- **Name:** Στην τελευταία στήλη αναγράφεται η ονομασία του μαθήματος.

Αυτά είναι όλα τα πεδία που θα δει κάποιος στη ΒΔ που θα χρησιμοποιηθεί για την οπτικοποίηση της πληροφορίας στην εφαρμογή που δημιουργήθηκε στο πλαίσιο αυτής της πτυχιακής εργασίας. Ο στόχος είναι αυτή η βάση να είναι ευέλικτη και να αναπροσαρμόζεται, και με κάθε νέα αξιολόγηση να εμπλουτίζεται με νέα δεδομένα, να βελτιώνεται και να παρέχει το υλικό για την οπτικοποίηση χρήσιμης πληροφορίας μεγαλύτερου όγκου.

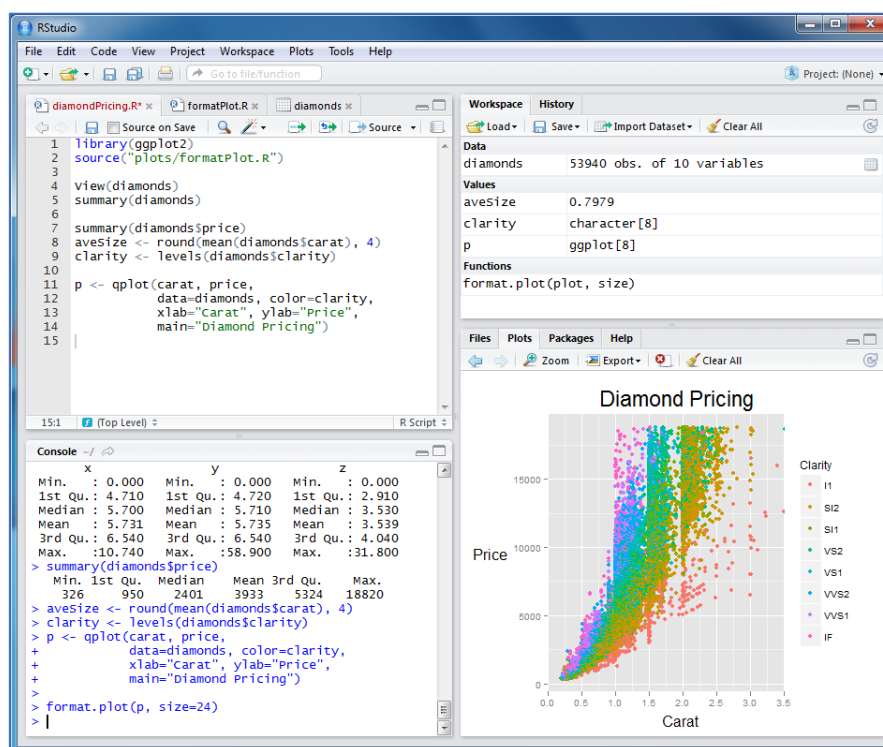
Στο επόμενο κεφάλαιο θα περιγραφούν τα εργαλεία και οι τεχνολογίες που χρησιμοποιήθηκαν τα οποία όχι απλά βοηθούν τους στόχους της εργασίας αλλά κατά μία έννοια, είναι ο λόγος ύπαρξης της.

Κεφάλαιο 4ο: Τεχνολογίες που Χρησιμοποιήθηκαν

Κατά την εκπόνηση της πτυχιακής εργασίας χρησιμοποιήθηκαν ποικίλα εργαλεία και τεχνολογίες που βοήθησαν στην επεξεργασία δεδομένων, και στον σχεδιασμό ενός λογισμικού για την οπτικοποίηση της πληροφορίας για τρίτους. Αυτές οι τεχνολογίες, οι οποίες θα αναλυθούν στο παρόν κεφάλαιο, συνδέονται άμεσα με την γλώσσα R είτε γιατί δημιουργήθηκαν με βάση την τεχνολογία της είτε γιατί δημιουργήθηκαν για να εξυπηρετούν λειτουργίες της και να τις επεκτείνουν.

4.1 RStudio IDE

Το RStudio, θεωρείται το πιο δημοφιλές εργαλείο σύνταξης κώδικα (στην πραγματικότητα μία εργαλειοθήκη) για τους χρήστες των τεχνολογιών που σχετίζονται με τη γλώσσα R και χαρακτηρίζεται ως ένα ολοκληρωμένο περιβάλλον ανάπτυξης (integrated development environment IDE) για τις γλώσσες R και Python [IS3].



Εικόνα 4.1 Το RStudio IDE [B12]

Το περιβάλλον χρήστη, όπως φαίνεται και στην Εικόνα 4.1, αποτελείται από μία κονσόλα στην οποία φαίνεται το αποτέλεσμα του κώδικα που εκτελεί ο χρήστης, από ένα παράθυρο σύνταξης και επεξεργασίας κώδικα και ένα περιβάλλον στο οποίο αναγράφονται τα στοιχεία που βρίσκονται στη μνήμη της εργαλειοθήκης. Το RStudio επιπλέον παρέχει εργαλεία για σχεδίαση, για ανάγνωση του ιστορικού και μία αποτελεσματική λειτουργία εντοπισμού σφαλμάτων (debugging). Ο χρήστης έχει τη δυνατότητα να επιλέξει όποια από τα παραπάνω στοιχεία θέλει να έχει εύκολα προσβάσιμα στην οθόνη του, και έτσι να διαχειριστεί τον χώρο εργασίας του (workspace management) όπως επιθυμεί.

Αυτή την ευελιξία έρχεται να συμπληρώσει το γεγονός ότι το RStudio είναι ανοικτού κώδικα που σημαίνει, εκτός των άλλων, ότι ο καθένας μπορεί να το εγκαταστήσει στην επιφάνεια εργασίας του έχοντας ως λειτουργικό σύστημα Windows, Mac ή Linux. [IS3].

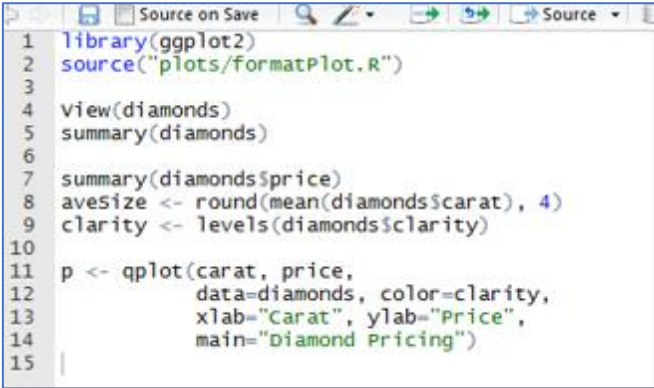
Για όλους του λόγους που αναφέρθηκαν, για την συγκεκριμένη εργασία έγινε αποκλειστική χρήση του RStudio ως το εργαλείο συγγραφής και ανάπτυξης κώδικα σε γλώσσα R. Ήδη για τις διαδικασίες που εκτελέστηκαν και αναφέρθηκαν στο δεύτερο κεφάλαιο, το εργαλείο σύνταξης κώδικα γλώσσας R ήταν αυτό το συγκεκριμένο.

4.2 Γλώσσα R

Η R, σε ένα γενικότερο πλαίσιο, χαρακτηρίζεται ως ένα ελεύθερο, ανοικτού κώδικα, περιβάλλον λογισμικού για στατιστικούς υπολογισμούς και γραφικά με χρήση υπολογιστή [IS4].

Αυτό το ελεύθερο περιβάλλον λογισμικού, εκτός των άλλων, διαθέτει δυνατότητα αποθήκευσης και διαχείρισης δεδομένων, μία ποικιλόμορφη και εύκολα κατανοητή συλλογή εργαλείων για την ανάλυση και εμφάνιση δεδομένων, και τέλος μία γλώσσα προγραμματισμού, την γλώσσα R [IS5].

Η γλώσσα R αναπτύχθηκε το 1993 από τον Ross Ihaka και τον Robert Gentleman και πρόκειται για μία καλά ανεπτυγμένη γλώσσα που χρησιμοποιείται συνήθως για στατιστική έρευνα, ανάλυση δεδομένων, δημιουργία γραφημάτων αλλά και για αλγόριθμους μηχανικής μάθησης. Η ίδια χαρακτηρίζεται ως καθολική καθώς είναι συμβατή με διαφορετικά περιβάλλοντα όπως τα Windows, τα Macintosh, τα UNIX, και τα Linux. Πολλοί οργανισμοί την επιλέγουν καθώς τα πλεονεκτήματα υπερτερούν των μειονεκτημάτων της [IS5].



```
1 library(ggplot2)
2 source("plots/formatPlot.R")
3
4 view(diamonds)
5 summary(diamonds)
6
7 summary(diamonds$price)
8 aveSize <- round(mean(diamonds$carat), 4)
9 clarity <- levels(diamonds$clarity)
10
11 p <- qplot(carat, price,
12           data=diamonds, color=clarity,
13           xlab="Carat", ylab="Price",
14           main="Diamond Pricing")
15
```

Εικόνα 4.2 Παράδειγμα κώδικα γλώσσας R.

4.2.1 Πλεονεκτήματα Γλώσσας R

Τα παρακάτω είναι κάποια από τα θετικά χαρακτηριστικά της γλώσσας R.

- Είναι διαθέσιμη για όλους καθώς είναι ανοιχτού κώδικα. Αυτό σημαίνει ότι δεν υπάρχουν χρεώσεις και ο καθένας μπορεί να την εγκαταστήσει στο τοπικό του μηχάνημα.
- Τρέχει ανεξάρτητα από την πλατφόρμα ή το λειτουργικό σύστημα που χρησιμοποιεί ο κάθε χρήστης.
- Διαθέτει διάφορα πακέτα και βιβλιοθήκες, κάποια από τα οποία θα αναφερθούν στη συνέχεια. Τα περισσότερα από αυτά τα πακέτα βρίσκονται στο CRAN, τον παγκόσμιο χώρο αποθήκευσης πακέτων ανοιχτού κώδικα.
- Είναι μία γλώσσα προσανατολισμένη στη στατιστική που την καθιστά ως ένα από τα κορυφαία εργαλεία για στατιστική και περαιτέρω ανάπτυξη στατιστικών εργαλείων. Γι'αυτό το λόγο πρωταγωνιστεί και στον κλάδο της Μηχανικής Μάθησης.

- Η γλώσσα R υποστηρίζεται από μία δραστήρια κοινότητα επιστημόνων η οποία ενισχύει την συνεχή αναβάθμιση και εξέλιξη της [AP1] [A2].

4.2.2 Μειονεκτήματα Γλώσσας R

Πέρα από τα θετικά στοιχεία της γλώσσας R που την καθορίζουν ως πρωταγωνίστρια στον χώρο της στατιστικής υπάρχουν και κάποια μειονεκτήματα που πρέπει να αναφερθούν.

- Είναι μία αρκετά περίπλοκη γλώσσα. Κάποιοι με ελάχιστη γνώση προγραμματισμού ίσως να δυσκολευτεί να την καταλάβει.
- Είναι πιο αργή στη μεταγλώττιση από άλλες γλώσσες προγραμματισμού, όπως η Python.
- Κατά πολλούς, είναι μία γλώσσα που δεν διαχειρίζεται σωστά τη μνήμη της. Βέβαια όταν χρησιμοποιούνται τα κατάλληλα εργαλεία και βιβλιοθήκες ανάλογα με τις ανάγκες του κάθε επιστήμονα αυτό το μειονέκτημα παύει να υφίσταται.
- Είναι μία γλώσσα ανοικτού κώδικα που υποστηρίζεται από το οικοσύστημα της R κάτι που μπορεί να οδηγήσει σε ασυνέπεια στην ποιότητα πακέτων, ειδικά αυτών που δεν προσφέρονται από το CRAN. Έτσι, θα πρέπει κάποιος να είναι προσεκτικός αλλά και έτοιμος να αναγνωρίσει τέτοιες ασυνέπειες.
- Η γλώσσα R δεν έχει αναπτυχθεί για δημιουργία ιστοσελίδων που απαιτείται η διαφύλαξη ευαίσθητης πληροφορίας από κακόβουλους χρήστες. Πλέον, έχουν δημιουργηθεί εργαλεία που αντιμετωπίζουν εν μέρει αυτό το πρόβλημα, κάποιο από τα οποία θα αναφερθεί στη συνέχεια [AP1] [A2].

4.3 Βιβλιοθήκες R

Όπως άλλες γλώσσες έτσι και η R διαθέτει βιβλιοθήκες και πακέτα. Ένα πακέτο R εμπεριέχει μία συλλογή μεθόδων (functions), κάποια μεταγλωττισμένα τμήματα κώδικα και κάποιες φορές δεδομένα. Τα πακέτα βοηθούν στην εύκολη και γρήγορη ανάπτυξη κώδικα. Κάθε πακέτο ανήκει σε μία βιβλιοθήκη η οποία καθορίζει τη θέση κάθε πακέτου στο σύστημα. Όταν συντάσσεται ένα τμήμα κώδικα γραμμένο σε γλώσσα R, στις πρώτες γραμμές συνιστάται να γίνεται η φόρτωση των βιβλιοθηκών και πακέτων όπως φαίνεται στην παρακάτω εικόνα [B3].

```
#install packages
install.packages("tidyr")

#load libraries
library(tidyr)
```

Εικόνα 4.3: Εγκατάσταση πακέτων και φόρτωση βιβλιοθηκών

Όπως προαναφέρθηκε, η επίσημη πηγή πακέτων ανοιχτού κώδικα είναι το CRAN στο οποίο μπορεί κάποιος να βρει σχεδόν 20.000 πακέτα τα οποία συνεχώς αυξάνονται [IS6]. Για τη συγκεκριμένη εργασία χρησιμοποιήθηκαν τουλάχιστον 12 πακέτα τα οποία είναι τα:

- **readxl**: Πακέτο που καθιστά εύκολη την ανάγνωση δεδομένων από Excel αρχεία.
- **xlsx**: Πακέτο που παρέχει λειτουργίες που καθιστούν δυνατή την ανάγνωση, εγγραφή και μορφοποίηση αρχείων Excel 2007 και Excel 97/2000/XP/2003.
- **dplyr**: Πακέτο με μεθόδους για τη διαχείριση δεδομένων κάνοντας χρήση αντικειμένων (objects).
- **plyr**: Πακέτο που παρέχει εργαλεία που λύνουν προβλήματα όπως ο διαμελισμός ενός μεγάλου προβλήματος σε μικρότερα διαχειρίσιμα κομμάτια τα οποία υφίστανται επεξεργασία από τον αναλυτή και ενώνονται πάλι.
- **reshape2**: Πακέτο που προσφέρει ευελιξία στην επεξεργασία και συγκέντρωση δεδομένων.

- **tidyr**: Πακέτο που προσφέρει εργαλεία για τη δημιουργία «τακτοποιημένων δεδομένων» δηλαδή δεδομένων που είναι εύκολο να διαχειριστείς και επεξεργαστείς.
- **stringr**: Πακέτο που προσφέρει συναρτήσεις για την διαχείριση τελεστών συμβολοσειρών (string operations) όπως τα στοιχεία «NAs».
- **ggplot2**: Πακέτο που παρέχει μεθόδους για τη δημιουργία αισθητικά όμορφων γραφημάτων.
- **shiny**: Ένα πακέτο που διευκολύνει τη δημιουργία διαδραστικών εφαρμογών στο περιβάλλον της R.
- **rmarkdown**: Πακέτο που βοηθάει στη δημιουργία δυναμικών εγγράφων/αναφορών που συνδυάζουν κώδικα, κείμενο, στατιστικές αναλύσεις με γραφήματα κ.α..
- **rconnect**: Πακέτο που δίνει τη δυνατότητα ανάπτυξης διεπαφής για έγγραφα R Markdown και εφαρμογές Shiny.
- **scales**: Πακέτο με εργαλεία για την οπτικοποίηση δεδομένων.

Στη συνέχεια θα περιγραφούν εκτενέστερα τα πακέτα Shiny και R Markdown καθώς αυτά χρησιμοποιήθηκαν σε μεγαλύτερο βαθμό.

4.4 Βιβλιοθήκη R Shiny

Το Shiny, είναι ένα πακέτο της R που ανήκει στην ομώνυμη βιβλιοθήκη. Το συγκεκριμένο πακέτο διευκολύνει τη δημιουργία διαδραστικών εφαρμογών για τον Ιστό και είναι ένα από τα βασικά εργαλεία που χρησιμοποιήθηκαν για τη δημιουργία της εφαρμογής γι' αυτή την πτυχιακή εργασία.

Με την R Shiny τεχνολογία δίνεται η δυνατότητα στον χρήστη να δημιουργήσει εφαρμογές που θα είναι προσβάσιμες από ιστοσελίδες ή να ενσωματώσει τις εφαρμογές σε έγγραφα R Markdown (αυτή η τεχνολογία θα αναλυθεί στη συνέχεια) και πολλά ακόμη. Επιπρόσθετα μία εφαρμογή Shiny μπορεί εύκολα να επεκταθεί αξιοποιώντας εργαλεία όπως τα θέματα CSS, τα htmlwidgets, την γλώσσα JavaScript, κ.α. [IS11]. Όταν ο χρήστης έχει ως στόχο την παρουσίαση δεδομένων και πληροφοριών με έναν εύκολο και φιλικό προς το κοινό τρόπο, τότε η R Shiny είναι μία επιλογή που καλύπτει τις ανάγκες του.

Ένα αρχείο κώδικα που περιέχει μια εφαρμογή Shiny έχει κατάληξη .R όπως και άλλα αρχεία κώδικα γραμμένα σε γλώσσα R. Ο κώδικας αποτελείται από δύο βασικά κομμάτια, το κομμάτι του διακομιστή (server) και το κομμάτι της διεπαφής χρήστη (user interface, UI), όπως φαίνεται και στην Εικόνα 4.4. Και τα δύο κομμάτια προγραμματίζονται σε γλώσσα R πράγμα που καθιστά εύκολη την εκτέλεση υπολογισμών, την επεξεργασία δεδομένων, και την οπτικοποίηση του αποτελέσματος μέσω της διεπαφής χρήστη [IS11].

```
library(shiny)

# Define UI ----
ui <- fluidPage(

)

# Define server logic ----
server <- function(input, output) {

}

# Run the app ----
shinyApp(ui = ui, server = server)
```

Εικόνα 4.4: Δομή ενός R Shiny εγγράφου κώδικα

Στο κομμάτι κώδικα της διεπαφής χρήστη πρέπει να προστεθεί ο κώδικας που αφορά την εικόνα και τα στοιχεία αλληλεπίδρασης ανθρώπου-μηχανής της εφαρμογής. Το πακέτο δίνει αρκετές επιλογές σχεδίασης μιας διεπαφής χρήστη μέσω των μεθόδων που προσφέρει [IS11]. Για την εργασία κάποιες από τις μεθόδους που χρησιμοποιήθηκαν και αφορούν την διεπαφή χρήστη είναι οι παρακάτω:

- **fluidPage():** Αυτή η μέθοδος χρησιμοποιείται για να δημιουργηθεί μία οθόνη η οποία προσαρμόζεται αυτόματα στις διαστάσεις του παραθύρου του προγράμματος περιήγησης του χρήστη.
- **titlePanel():** Αυτή η μέθοδος χρησιμοποιείται από την fluidPage() μέθοδο και ορίζει τον τίτλο της σελίδας.
- **tabsetPanel():** Με αυτή τη μέθοδο ο προγραμματιστής μπορεί να δημιουργήσει διαφορετικές καρτέλες στην ίδια σελίδα, δημιουργώντας την αίσθηση καταλόγου.
- **fileInput():** Αυτή η μέθοδος προσθέτει ένα στοιχείο στην σελίδα που επιτρέπει την φόρτωση αρχείων κατά τη χρήση της εφαρμογής Shiny.
- **selectInput():** Μία από τις πιο συχνά χρησιμοποιούμενες μεθόδους κατά την οποία μπορεί κάποιος να προσθέσει ένα 'select' στοιχείο στην σελίδα του.
- **downloadButton():** Αυτή η μέθοδος χρησιμοποιείται ώστε να δοθεί η δυνατότητα λήψης αρχείου που έχει δημιουργηθεί κατά την χρήση της εφαρμογής.

Οι μέθοδοι και οι δυνατότητες που παρέχει το πακέτο Shiny είναι πολλές και οι επιλογές συνεχώς αυξάνονται. Το ίδιο πλούσιες είναι και οι δυνατότητες του πακέτου R Markdown.

4.5 Βιβλιοθήκη R Markdown

Η τεχνολογία R Markdown είναι για πολλούς το εργαλείο με το οποίο ένας αναλυτής μοιράζεται την πληροφορία που θέλει με έναν συνεργάτη του, στην ομάδα του ή σε ένα τρίτο πρόσωπο μέσω δυναμικών αναφορών (dynamic reports). Στόχος του συγκεκριμένου εργαλείου είναι να παρέχει στον αναλυτή έναν εύκολο και ευέλικτο τρόπο οργάνωσης και παρουσίασης των δεδομένων και της πληροφορίας που θέλει να διαμοιράσει. Να παρέχει, έναν εύκολο τρόπο ώστε να μπορεί να ανανεώνει την πληροφορία όταν χρειάζεται χωρίς πολλά βήματα, και έναν ευέλικτο τρόπο ώστε να παρέχει την πληροφορία σε διαφορετική μορφή κάθε φορά, ανάλογα με τις ανάγκες της κάθε περίπτωσης. Αυτό συμβαίνει μέσω ενός περιβάλλοντος ανάπτυξης κώδικα και κειμένου, που προσφέρει το πακέτο R Markdown, όπου η πληροφορία παραμένει προσβάσιμη, ακριβής και συνεπής, με δυνατότητα χρήσης άλλων τεχνολογιών και γλωσσών πέραν όσων παρέχει η R [B4].

4.5.1 Τα Βασικά Σημεία της R Markdown

Η R Markdown είναι ένα πακέτο της R το οποίο δημιουργήθηκε το 2014 και τα τελευταία χρόνια σταδιακά εξελίχθηκε σε ένα ολοκληρωμένο σύστημα για την σύνταξη εγγράφων [B4]. Τα αρχεία που αφορούν αυτό το πακέτο έχουν την κατάληξη .Rmd, και έτσι διακρίνονται από άλλα αρχεία .R.

Μέσα από αυτό το πακέτο δίνεται η δυνατότητα για δημιουργία δυναμικών αρχείων και αναφορών κάνοντας χρήση απλού συντακτικού σε συνδυασμό με κομμάτια κώδικα γλωσσών προγραμματισμού, όταν αυτό χρειαστεί. Το χαρακτηριστικό αυτό είναι σημαντικό για την επιστήμη των δεδομένων, καθώς επιτρέπει στον ερευνητή την ενσωμάτωση στοιχείων εμπλουτισμένου κειμένου σε έγγραφα απλού κειμένου, ενώ ταυτόχρονα αυτά τα έγγραφα μπορούν να συνδεθούν με δεδομένα και να εκτελούν κώδικα [IS7].

Η ευελιξία αυτού του συστήματος σύνταξης επιτρέπει στον προγραμματιστή να μεταμορφώνει το αρχικό αρχείο κώδικα (με την κατάληξη .Rmd) σε διάφορους τύπους αρχείων, και τελικά να

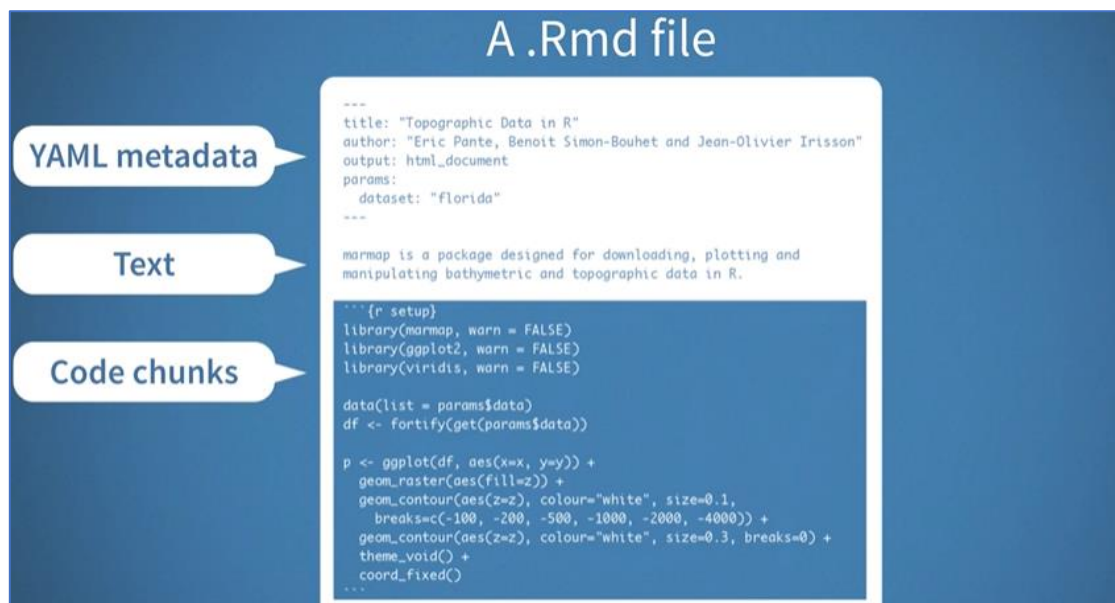
δημιουργήσει τις αναφορές που επιθυμεί. Αυτές οι αναφορές μπορεί να είναι μορφής αρχείων MS Word, αρχείων PDF, μία ολοκληρωμένη ιστοσελίδα κ.α. ώστε να αναπαράγεται η πληροφορία με διάφορους τρόπους [IS7]. Ακόμη, αν ο ερευνητής θελήσει να αλλάξει κάτι στο αρχείο PDF ή στο αρχείο MS Word μπορεί απλώς να διορθώσει ή να ενημερώσει τον πηγαίο κώδικα στο .Rmd αρχείο και να το μεταγλωττίσει ξανά σε όποια μορφή θέλει.

4.5.2 Η Δομή και το Συντακτικό Ενός Αρχείου R Markdown

Ένα αρχείο κώδικα R Markdown έχει πάντα την κατάληξη .Rmd και αποτελείται από τριών ειδών τύπους περιεχομένου.

Όπως φαίνεται στην Εικόνα 4.1, αρχικά υπάρχει μία προαιρετική επικεφαλίδα γραμμένη σε γλώσσα YAML (Ain't Markup Language). Στην επικεφαλίδα δηλώνονται όλα τα μεταδεδομένα (metadata) τα οποία καθορίζουν σημαντικά στοιχεία για την ανάπτυξη του κώδικα στο αρχείο, ενώ δηλώνονται επίσης και οι παράμετροι που πρέπει να είναι προσβάσιμοι κάθε στιγμή από το πρόγραμμα. Στη συνέχεια προστίθενται τα τμήματα κώδικα (code chunks) τα οποία εκτελούνται όταν το αρχείο μεταγλωττιστεί. Τέλος, προστίθεται το κείμενο (text) που συμπληρώνει το περιεχόμενο του αρχείου. Τα δύο τελευταία κομμάτια περιεχομένου, τα τμήματα κώδικα και το κείμενο μπορούν να τοποθετηθούν με διαφορετική σειρά, η οποία καθορίζεται από τον προγραμματιστή σύμφωνα με το τελικό αποτέλεσμα που θέλει να έχει [B4].

Το κείμενο είναι γραμμένο σε γλώσσα Markdown η οποία χαρακτηρίζεται ως μία εξαιρετικά φιλική προς τον χρήστη γλώσσα σύνταξης με πολλές επιλογές για την διαμόρφωση κειμένου [IS8].



Εικόνα 4.5: Η βασική δομή ενός αρχείου R Markdown

Τα τμήματα κώδικα μπορούν να εμπεριέχουν παράλληλα τμήματα κώδικα από διάφορες γλώσσες όπως R, Python, SQL, Javascript, CSS, C, Fortran, κ.α.. Η μεταγλώττιση αυτού του κώδικα γίνεται δυνατή μέσω του πακέτου knitr το οποίο είναι ένα πακέτο γενικού σκοπού για τη δημιουργία αναφορών [IS6]. Έτσι, ο προγραμματιστής μπορεί να συντάξει τον κώδικα που επιθυμεί, να τον εκτελέσει, και να αποθηκεύσει το αποτέλεσμα σε οποιαδήποτε μορφή θέλει ο ίδιος και η ομάδα του [B4].

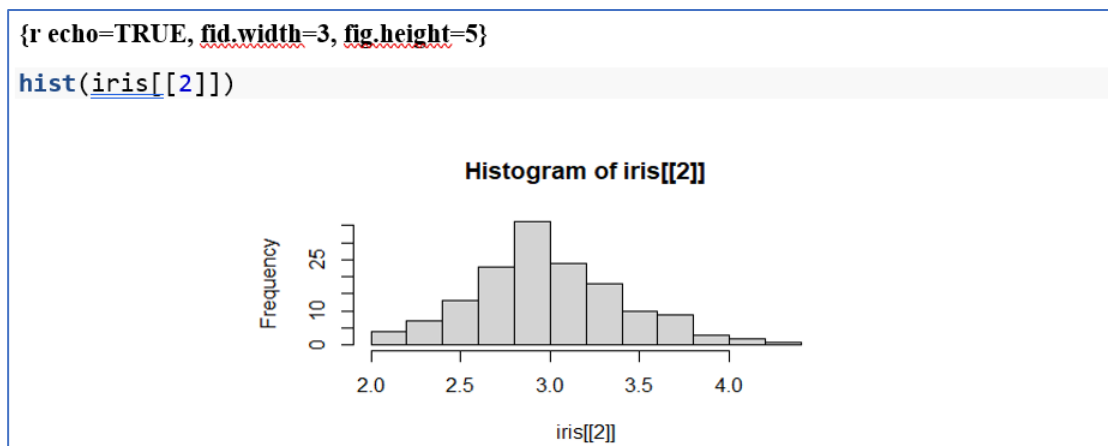
Στη συνέχεια της παραγράφου δίνονται κάποια παραδείγματα εντολών που αφορούν την προσθήκη γλώσσας προγραμματισμού σ' ένα έγγραφο R Markdown [B4].

- **Προσθήκη τμημάτων κώδικα διαφορετικών γλωσσών σε ένα αρχείο κώδικα R Markdown:** Όταν ο προγραμματιστής επιθυμεί να προσθέσει στο αρχείο R Markdown μία γλώσσα προγραμματισμού πέρα του απλού ή εμπλουτισμένου κειμένου, τότε μπορεί να το κάνει είτε μέσω προσθήκης τμημάτων κώδικα, γνωστά ως code chunks, όπως φαίνεται στην Εικόνα 4.5, ή διαφορετικά μπορεί να προσθέσει το τμήμα κώδικα μέσα στο κείμενο του με την χρήση της αποστρόφου «`» όπως φαίνεται στο παράδειγμα ‘`r Sys.Date()`. Στο συγκεκριμένο παράδειγμα, η γλώσσα προγραμματισμού δηλώνεται στην αρχή της εντολής ‘r’ και στη συνέχεια τοποθετείται η εντολή που επιθυμεί ο προγραμματιστής ‘Sys.Date()’.

```
```{r}
some code
```
```

Εικόνα 4.5: Προσθήκη κώδικα R σε ένα αρχείο κώδικα R Markdown

- **Η εντολή {r echo=TRUE ή FALSE}:** Με αυτή την εντολή ο αναγνώστης του παραγόμενου αρχείου βλέπει ή δεν βλέπει τον κώδικα από τον οποίο προήλθε το περιεχόμενο του αρχείου. Για παράδειγμα στην Εικόνα 4.6 όπου {r echo=TRUE} ο αναγνώστης μπορεί να δει εκτός του ιστογράμματος και το τμήμα κώδικα «hist(iris[[2]])» στην αναφορά που έχει παραχθεί. Οι μεταβλητές `fig.width` και `fig.height` καθορίζουν τις διαστάσεις του γραφήματος.



Εικόνα 4.6: Παράδειγμα χρήσης της εντολής ‘echo’


- **Η εντολή {r eval= TRUE ή FALSE }:** Σε ορισμένες περιπτώσεις ο προγραμματιστής επιθυμεί να εμφανίσει τον κώδικά του και όχι να τον εκτελέσει. Η εντολή ‘eval’ προσφέρει αυτή τη δυνατότητα εφόσον οριστεί ως ‘TRUE’.

Τα παραπάνω παραδείγματα είναι μερικές από τις εντολές που χρησιμοποιήθηκαν στη συγκεκριμένη εργασία για την παραγωγή αναφορών.

4.5.3 Η Γλώσσα Markdown

Το κείμενο σε ένα αρχείο R Markdown γράφεται σε γλώσσα σύνταξης Markdown. Η γλώσσα Markdown σχεδιάστηκε με κύριο γνώμονα την ευκολία χρήσης έτσι ώστε να είναι προσιτή και

κατανοητή από άτομα που δεν έχουν απαραίτητα εμπειρία με τον προγραμματισμό. Κάποια από τα βασικά στοιχεία σύνταξης κειμένου παρουσιάζονται στον παρακάτω πίνακα ενώ η πλήρης λίστα των στοιχείων μπορεί να βρεθεί στην επίσημη σελίδα της γλώσσας [IS9].

| Syntax | Becomes |
|---|--|
| Plain text | Plain text |
| <i>*italics*</i> or <code>_italics_</code> | <i>italics</i> |
| **Bold** or <code>__Bold__</code> | Bold |
| --strikethrough-- | –strikethrough– |
| <code>[link](https://www.ihu.gr/)</code> | link |
| inline equation: <code>\$A = \pi*r^{2}\$</code> | $A = \pi * r^2$ |
| image: <code></code> | image:  |

Πίνακας 4.1: Βασικά στοιχεία σύνταξης κειμένου της γλώσσας Markdown

4.5.4 Τα Παραγόμενα Αρχεία

Από ένα αρχείο `.Rmd` ένας αναλυτής μπορεί να δημιουργήσει τις αναφορές που θέλει σε μορφές αρχείων που αυτός επιθυμεί, όπως ένα αρχείο MS Word, ένα αρχείο PDF, μία ιστοσελίδα ή ακόμη και ένα βιβλίο.

Πρακτικά, αυτό γίνεται μέσω της εντολής `render` ή της χρήσης του κουμπιού `Knit` μέσα από το περιβάλλον χρήσης του Rstudio. Πιο αναλυτικά, όταν κάποιος τρέξει την εντολή `render`, για το `.Rmd` αρχείο, το σύστημα κάνει χρήση του πακέτου `knitr`, το οποίο χαρακτηρίζεται ως μία μηχανή δυναμικής δημιουργίας αναφορών κάνοντας χρήση της γλώσσας R [B4]. Η μηχανή εκτελεί όλα τα κομμάτια κώδικα και δημιουργεί ένα νέο αρχείο `markdown (.md)` το οποίο εμπεριέχει τον κώδικα και το αποτέλεσμα αυτού. Στη συνέχεια το αρχείο `.md` υφίσταται επεξεργασία από το Pandoc, που είναι ένας γενικός μετατροπέας εγγράφων, και δημιουργείται η τελική μορφή της αναφοράς [IS9].

4.6 Shinyapps.io

Μία εφαρμογή που βασίζεται στην τεχνολογία του πακέτου Shiny, για να είναι διαθέσιμη διαδικτυακά, χρειάζεται να τοποθετηθεί σε κάποιον διακομιστή (server). Υπάρχουν ποικίλοι διακομιστές που μπορούν να φιλοξενήσουν εφαρμογές Shiny. Στο πλαίσιο της πτυχιακής εργασίας αξιοποιήθηκε ο διακομιστής που παρέχεται από το περιβάλλον Shinyapps.io. Ο διακομιστής της πλατφόρμας παρέχεται από την «Amazon’s Web Services (AWS)» [IS11].

Η Shinyapps.io είναι μια διαδικτυακή πλατφόρμα η οποία δημιουργήθηκε από το R Studio, και μέσα από αυτή ο αναλυτής μπορεί να μοιραστεί τις εφαρμογές που έχει δημιουργήσει. Κάθε εφαρμογή που φιλοξενείται στον διακομιστή είναι αυτόνομη και λειτουργεί είτε με δεδομένα που έχουν φορτωθεί στον διακομιστή από τον ερευνητή, είτε με δεδομένα που αντλούνται από βιβλιοθήκες κατά την χρήση της εφαρμογής.

Για να γίνει χρήση της πλατφόρμας πρέπει αρχικά να δημιουργηθεί ένας λογαριασμός χρήστη. Ο λογαριασμός μπορεί να διαθέτει απλά τις απαραίτητες λειτουργίες και να έχει μηδενικό κόστος, κάτι που αφορά άτομα που διαχειρίζονται μικρής έκτασης έργα ή εργασίες ή μπορεί να προσφέρει μία λίστα

από υπηρεσίες, με επιπλέον κόστος [B2]. Η δεύτερη κατηγορία προϊόντων αφορά κυρίως μεγαλύτερους οργανισμούς και έργα. Τα κύρια χαρακτηριστικά που οδήγησαν στην χρήση του Shinyapps.io και αναλύονται παρακάτω είναι η ευχρηστία η ασφάλεια και η ευελιξία.

4.6.1 Ευχρηστία

Μετά από μελέτη και δοκιμές διαφόρων πλατφορμών και διακομιστών έγινε η χρήση της πλατφόρμας Shinyapps.io. Καταλύτης σε αυτή την απόφαση ήταν η ευκολία εγκατάστασης του εργαλείου καθώς και η γενικότερη χρήση του, δεδομένου ότι προσφέρει ένα πολύ εύχρηστο περιβάλλον λειτουργίας, όπου μεταξύ άλλων παρακολουθείται ο αριθμός ατόμων που την επισκέφτηκαν.

Εφόσον κάποιος επιλέξει τη δωρεάν συνδρομή της πλατφόρμας έχει τις υπηρεσίες που του προσφέρονται για έως και πέντε εφαρμογές. Ο συνολικός χρόνος ενεργών ωρών κάθε εφαρμογής είναι εικοσιπέντε (25) ώρες ανά μήνα. Για τις ανάγκες της πτυχιακής εργασίας χρησιμοποιήθηκε η δωρεάν διάθεση της υπηρεσίας καθώς η εφαρμογή είναι μοναδική και οι ώρες χρήσης της εφαρμογής είναι συνήθως κάτω από δέκα (10) ώρες ανά μήνα [IS12].

4.6.2 Ασφάλεια

Στην αρχή του κεφαλαίου είχε αναφερθεί ότι η γλώσσα R δεν είναι ασφαλής για δημιουργία ιστοσελίδων. Αυτό αντιμετωπίζεται σε μεγάλο βαθμό κάνοντας χρήση της πλατφόρμας shinyapps.io η οποία έχει σχεδιαστεί με γνώμονα την ασφάλεια των δεδομένων (secure-by-design). Κάθε εφαρμογή εκτελείται στο δικό της προστατευμένο περιβάλλον και η πρόσβαση είναι πάντα κρυπτογραφημένη με το πρωτόκολλο ασφαλείας SSL [B6]. Όλα τα πλάνα συμμετοχής, είτε με κόστος συνδρομής είτε χωρίς κόστος, παρέχουν έλεγχο ταυτότητας χρήστη, αποτρέποντας ανώνυμους χρήστες να έχουν πρόσβαση στις εφαρμογές.

4.6.3 Ευελιξία

Η πλατφόρμα shinyapps.io επιτρέπει στον χρήστη να εξελίξει την εργασίας του, να αλλάξει την εφαρμογή του όσες φορές θεωρεί αναγκαίο αλλά και να προσθέσει νέες εφαρμογές όποτε αυτός επιθυμεί, χωρίς επιπλέον κόστος [IS12].

Κεφάλαιο 5ο: Οπτικοποίηση Δεδομένων Ερωτηματολογίου

Στα προηγούμενα κεφάλαια έγινε αναλυτική περιγραφή στα χαρακτηριστικά των πρωτογενών δεδομένων που χρησιμοποιήθηκαν, στην επεξεργασία και τον μετασχηματισμό ένταξης τους σε μια ολοκληρωμένη βάση δεδομένων, καθώς και σε όλες τις τεχνολογίες που χρησιμοποιήθηκαν σε κάθε μία επιμέρους διαδικασία. Στην παρούσα ενότητα περιγράφεται η οπτικοποίηση των δεδομένων, που ακολουθεί τον μετασχηματισμό τους.

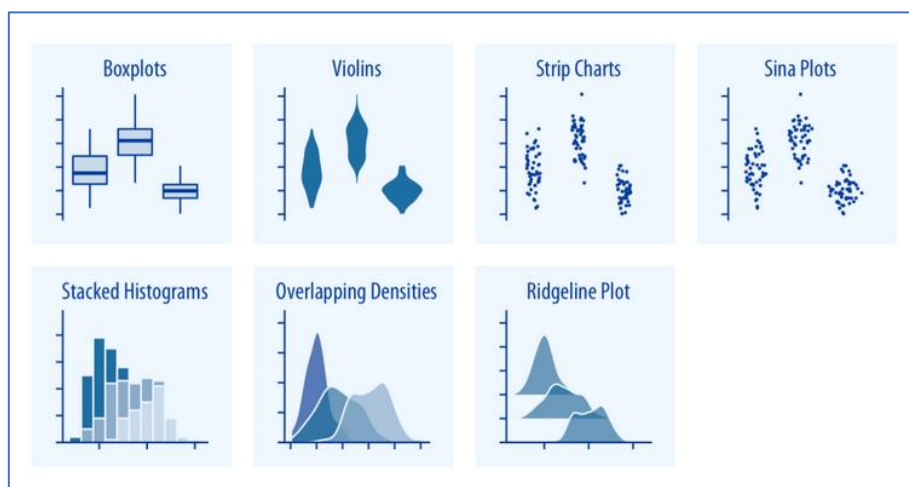
Η οπτικοποίηση των δεδομένων βοηθά στην κατανόηση τους μέσω της αναπαράστασης αυτών σε γραφήματα. Ένας πίνακας με δεδομένα μπορεί μέσω τεχνικών οπτικοποίησης να προσφέρει πολύ καλύτερη απεικόνιση της πληροφορίας καθώς τα γραφήματα βοηθούν τον αναγνώστη να αναγνωρίζει πιο εύκολα μοτίβα και να αντλεί αποτελεσματικότερα τη χρήσιμη πληροφορία.

5.1 Οπτικοποίηση Δεδομένων

Ο Claus Wilke στο βιβλίο του «Fundamentals of Data Visualization A Primer on Making Informative and Compelling Figures» αναφέρει χαρακτηριστικά ότι η οπτικοποίηση δεδομένων είναι ένας κλάδος της Επιστήμης των Δεδομένων που συνδυάζει μ' έναν ιδιαίτερο τρόπο την Τέχνη και την Επιστήμη [B11].

Βασικό πυλώνα αυτής της σχέσης αποτελεί ο συνδυασμός μίας αισθητικά όμορφης αποτύπωσης της εικόνας (Τέχνη) με την ορθή και ακριβή χρήση των δεδομένων (Επιστήμη). Ένα γράφημα δεν πρέπει να παραπλανά τον αναγνώστη, ενώ ταυτόχρονα χρειάζεται να είναι αισθητικά ευχάριστο με ευδιάκριτη την πληροφορία, καθώς οι αισθητικές προσθήκες πρέπει να τείνουν να ενισχύουν το μήνυμα που θέλει να περάσει ο ερευνητής [B11].

Όταν ένας ερευνητής επιθυμεί να οπτικοποιήσει δεδομένα, αναλύει τις τιμές των δεδομένων και τις μετατρέπει μ' έναν λογικό τρόπο σε οπτικά στοιχεία που συνθέτουν ένα γράφημα. Στον κόσμο της οπτικοποίησης δεδομένων υπάρχουν πολλοί διαφορετικοί τύποι γραφημάτων, για παράδειγμα το γράφημα γραμμής, το κυκλικό διάγραμμα, το διάγραμμα διασποράς, ο χάρτης θερμοτήτων, κ.α. Η ίδια πληροφορία μπορεί να αποτυπωθεί με διαφορετικούς τρόπους ανάλογα με τις ανάγκες του κάθε έργου. Με μία πρώτη όψη, στην Εικόνα 5.1, τα γραφήματα δεν φαίνεται να έχουν κοινά, και όμως όλες οι απεικονίσεις, όταν μελετηθούν προσεκτικά φαίνεται πως αποτυπώνουν την ίδια πληροφορία καθώς οι τιμές των δεδομένων και ο αλγόριθμος επεξεργασίας τους είναι ίδιοι [B11].



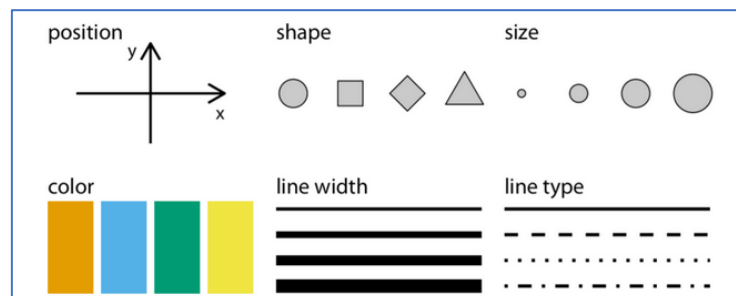
Εικόνα 5.1: Παραδείγματα γραφημάτων

Σε αυτή την πτυχιακή εργασία η οπτικοποίηση των δεδομένων βασίζεται σε δύο τύπους γραφημάτων, το ραβδόγραμμα και το γράφημα πυκνότητας. Αυτοί οι δύο τύποι χαρακτηρίζονται για την υψηλή αισθητική τους αλλά και την απλότητα τους. Πριν περιγραφούν πιο αναλυτικά οι συγκεκριμένοι τύποι γραφημάτων θα δοθεί ο ορισμός της Γραμματικής των Γραφημάτων.

5.2 Η Γραμματική των Γραφημάτων

Ο Leland Wilkinson στην «Γραμματική των Γραφημάτων» [B13] υποστηρίζει ότι τα περισσότερα γραφήματα κατασκευάζονται μέσω ενός πλαισίου δεδομένων (data frame), το οποίο είναι ο πιο συνηθισμένος τρόπος αποθήκευσης δεδομένων στην R αλλά και γενικότερα στην ανάλυση δεδομένων [IS13], και ότι αποτελούνται από κάποια βασικά στοιχεία. Τα χαρακτηριστικά αυτά στοιχεία που περιγράφουν κάθε γράφημα και κατ' επέκταση τα γραφήματα της συγκεκριμένης εργασίας είναι τα εξής [A5]:

- **Δεδομένα (data):** Η βάση κάθε γραφήματος είναι τα δεδομένα που ο κάθε επιστήμονας ή ομάδα επιστημόνων θέλουν να επεξεργαστούν και να οπτικοποιήσουν μέσω γραφημάτων.
- **Γεωμετρικά αντικείμενα ή γλύφοι (geometric objects ή glyphs):** Οι βασικές μονάδες ενός γραφήματος. Τα γεωμετρικά αντικείμενα αντιπροσωπεύουν μία ή περισσότερες μοναδικές περιπτώσεις στο πλαίσιο δεδομένων. Στο γράφημα μπορεί να πάρουν την μορφή γραμμής, σημείου, ορθογωνίου και άλλων σχημάτων.
- **Αισθητικές (aesthetics):** Οι διαφορετικές ιδιότητες των γεωμετρικών αντικειμένων οι οποίες ποικίλλουν ανάλογα με τις τιμές των μεταβλητών, της περίπτωσης που αντιπροσωπεύει το κάθε γεωμετρικό αντικείμενο. Ο κατάλογος ιδιοτήτων ενός γεωμετρικού αντικειμένου περιλαμβάνει τη θέση, το σχήμα, το μέγεθος, το χρώμα κ.α.. Βάσει της «Γραμματικής των Γραφημάτων», ανάλογα με το γράφημα που χρησιμοποιείται, κάποιες από αυτές τις ιδιότητες σημειώνονται ως Αισθητικές και χρησιμοποιούνται για την αναπαράσταση της πληροφορίας και κάποιες όχι. Παραδείγματος χάριν, σε ένα διάγραμμα όπου μέρος της πληροφορίας συμβολίζεται με σημεία (τελείες), η ιδιότητα του μεγέθους αλλά και του σχήματος δε θεωρούνται Αισθητικές καθώς όλα τα σημεία έχουν το ίδιο μέγεθος και το ίδιο σχήμα. Στο ίδιο παράδειγμα θα μπορούσε να χρησιμοποιηθεί ως μία Αισθητική το χρώμα. Κάθε σημείο, ανάλογα με την τιμή της ιδιότητας που έχει, θα μπορούσε να ανήκει σε μία κατηγορία η οποία συμβολίζεται με διαφορετικό χρώμα. Μία κατηγοριοποίηση που συναντάται συχνά είναι η ηλικιακή κατηγοριοποίηση όπου μία ηλικιακή ομάδα αντιστοιχίζεται με κάποιο χρώμα, μία δεύτερη ηλικιακή ομάδα με διαφορετικό χρώμα και μία τρίτη ομάδα με ένα άλλο χρώμα. Έτσι, κάθε σημείο του γραφήματος έχει χρώμα ανάλογο της ηλικιακής ομάδας του δείγματος.



Εικόνα 5.2: Παραδείγματα στοιχείων γεωμετρικών αντικειμένων και Αισθητικών

- **Πλαίσιο (frame):** Συγκεκριμένη ιδιότητα των Αισθητικών που συσχετίζει τη θέση κάθε γλύφου (μορφή συμβόλου) στο γράφημα με τιμές μεταβλητών για τις περιπτώσεις που αντιπροσωπεύει ο κάθε γλύφος.

- **Κλίμακες (scales):** Οι κλίμακες είναι μία από τις ιδιότητες των Αισθητικών και περιγράφονται ως επιλογές που καθορίζουν την ακριβή σχέση μεταξύ ιδιοτήτων των Αισθητικών και των τιμών δεδομένων για τα γεωμετρικά αντικείμενα. Χρησιμοποιώντας το παράδειγμα που αναφέρθηκε, μέσω των Κλιμάκων δηλώνεται η πολιτική αντιστοίχισης των χρωμάτων βάσει της ηλικιακής ομάδας. Πρακτικά, με τις Κλίμακες δηλώνεται ποια ηλικιακή ομάδα αντιστοιχίζεται σε ποιο χρώμα.
- **Στατιστικά (statistics):** Τα περισσότερα διαγράμματα αναπαριστούν στατιστικές πληροφορίες σχετικές με δεδομένα, όπως ο Μ.Ο., το πλήθος, η διακύμανση έτσι ώστε να κατανοηθεί καλύτερα η πληροφορία, με τρόπο απλό και εποπτικό [A5].

```

ggplot(df2_subset(), aes(QN,value,fill=variable ))+
  geom_bar(stat="identity",position="dodge")+
  xlab("ΕΡΩΤΗΣΗ") +
  ylab("ΒΑΘΜΟΛΟΓΙΑ") +
  ggtitle( paste0(
    "ΔΙΠΑΕ, Τμήμα Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων \n"
    ,"Ακαδημαϊκό έτος ", input$Year2, ' ', semester, ' Εξάμηνο', '\n', input$Course2)) +
  geom_text(aes(label = value2), colour = "black",hjust=-0.2, vjust=0.5,
    position = position_dodge(0.9), angle= 90, size=2.9 )+
  scale_fill_discrete(labels = c(plithos_A_counter,plithos_C_counter)) +
  theme(
    axis.text.x = element_text(angle = -70, vjust = 0.0, hjust=0),
    axis.title.x = element_text(color="#797473", size=12, face="bold"),
    axis.title.y = element_text(color="#797473", size=12, face="bold"),
    axis.line = element_line(colour = "grey", size = 1, linetype = "solid"),
    legend.title = element_blank(),
    plot.title = element_text(hjust = 0.5)
  )

```

Εικόνα 5.3: Απόσπασμα κώδικα της πτυχιακής εργασίας που περιέχει τα θεμελιώδη στοιχεία της γραμματικής των γραφημάτων σε γλώσσα R

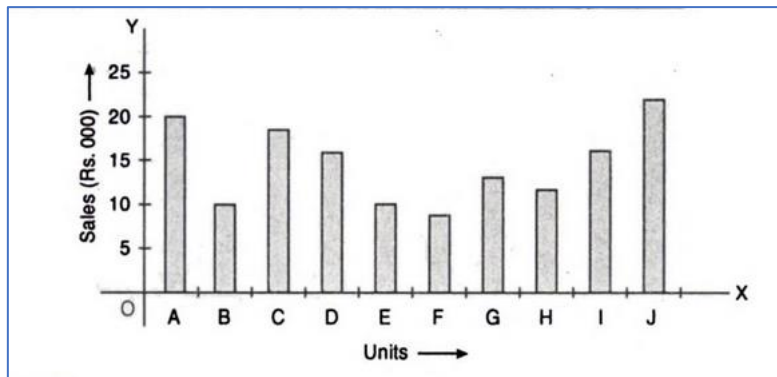
Στην περίπτωση της πτυχιακής εργασίας η εφαρμογή μέσω της οποίας γίνεται η οπτικοποίηση των δεδομένων δέχεται τη ΒΔ η οποία είναι τύπου αρχείου .rds. και τη μετατρέπει, μέσω κώδικα R, σε ένα πλαίσιο δεδομένων (data frame).

Το απόσπασμα κώδικα της Εικόνας 5.3 αφορά την παραγωγή γραφήματος τύπου ραβδόγραμματος. Το ραβδόγραμμα είναι ένας από τους δύο βασικούς τύπους γραφημάτων που χρησιμοποιήθηκαν για την οπτικοποίηση των δεδομένων για την πτυχιακή εργασία.

5.3 Ραβδόγραμμα Αξιολογήσεων

Το ραβδόγραμμα (Εικόνα 5.4) είναι το πιο συχνά χρησιμοποιούμενο γράφημα για την παρουσίαση και οπτικοποίηση δεδομένων και χρησιμοποιείται σε διάφορα επιστημονικά πεδία όπως η ιατρική, η βιολογία, η χρηματοοικονομική κ.α..

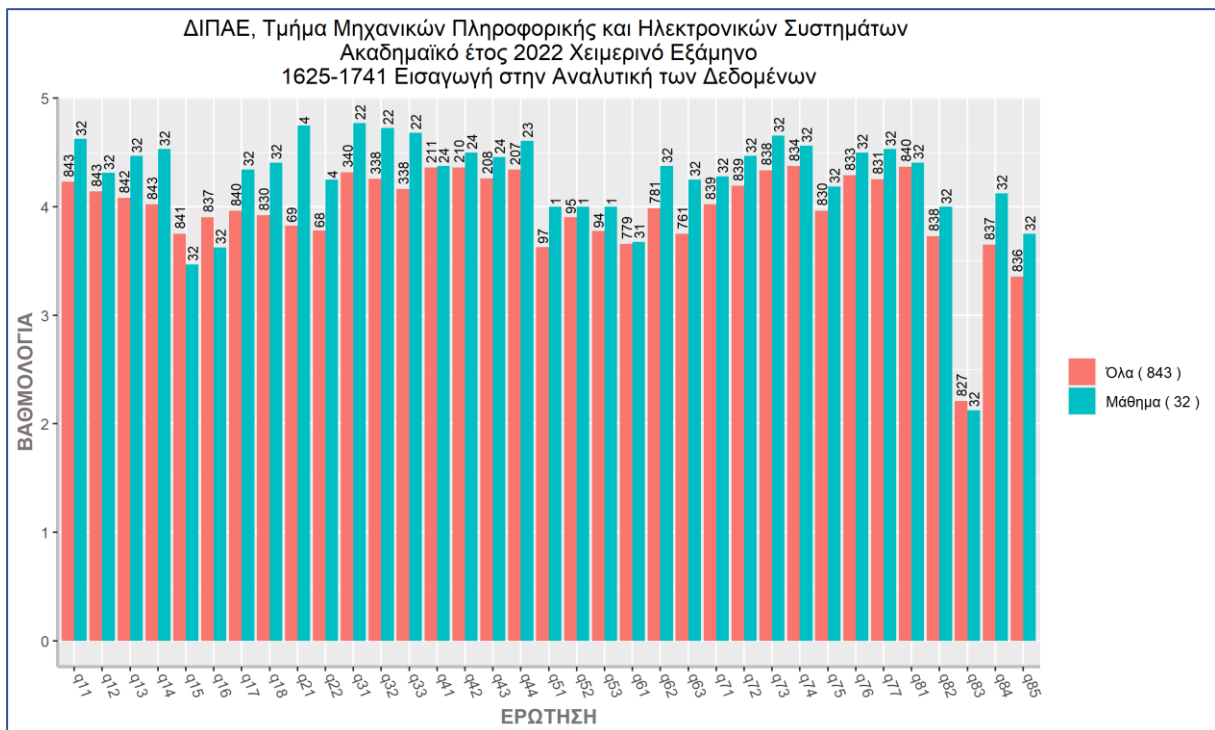
Το ραβδόγραμμα ορίζεται ως ένα γράφημα το οποίο αποτελείται από μια ομάδα ορθογωνίων ίσων αποστάσεων που αποτυπώνουν τιμές μεταβλητών. Είναι μονοδιάστατα διαγράμματα, επομένως το ύψος αναλογεί στις τιμές των μεταβλητών ενώ το πλάτος δεν αποτυπώνει κάποια πληροφορία. Το ραβδόγραμμα είναι μία τεχνική οπτικοποίησης η οποία είναι απλή στην κατασκευή και ερμηνεύεται εύκολα από τον αναγνώστη, ως εκ τούτου η πληροφορία μπορεί εύκολα να αξιοποιηθεί για την εξαγωγή συμπερασμάτων [B7] [B10].



Εικόνα 5.4: Το ραβδόγραμμα

Ένας από τους στόχους οπτικοποίησης που τέθηκαν κατά την έναρξη της πτυχιακής εργασίας ήταν η αποτύπωση της σύγκρισης δύο μέσων όρων (Μ.Ο.). Πιο συγκεκριμένα τον Μ.Ο. των τιμών από τις απαντήσεις των ερωτήσεων αξιολόγησης ενός μαθήματος, σε σχέση με τον Μ.Ο. των τιμών των απαντησεων του συνόλου των μαθημάτων που συμπεριλαμβάνονται στο Π.Σ. του εν λόγω μαθήματος. Αυτή η πληροφορία θα βοηθήσει το ακαδημαϊκό προσωπικό να αντιληφθεί πως βαθμολογείται το κάθε μάθημα σε σύγκριση με τα υπόλοιπα μαθήματα του ίδιου Π.Σ..

Η πρώτη τεχνική οπτικοποίησης δεδομένων που χρησιμοποιήθηκε στο πλαίσιο της εργασίας είναι το γράφημα πολλαπλών ράβδων, που αποτελεί υποκατηγορία του τύπου ραβδογράμματος. Μέσω αυτού του τύπου γραφήματος μπορεί να αποτυπωθεί αποτελεσματικά η απόδοση κάθε μαθήματος σε σύγκριση με το σύνολο των μαθημάτων όπως φαίνεται και στην Εικόνα 5.5.



Εικόνα 5.5 Διάγραμμα πολλαπλών ράβδων για την σύγκριση Μ.Ο. τιμών των απαντήσεων ερωτηματολογίου ενός μαθήματος με το σύνολο των μαθημάτων .

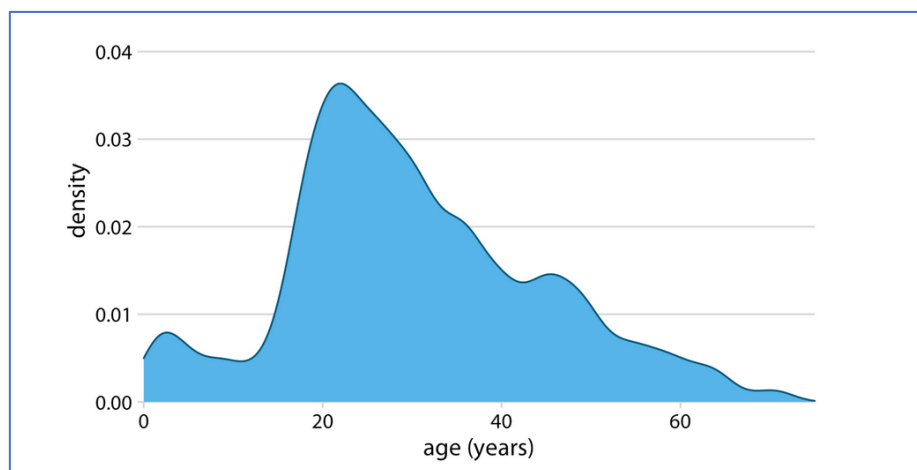
Στην παρουσίαση της σύγκρισης των Μ.Ο. των τιμών μέσω του ραβδογράμματος (Εικόνα 5.5) παρατηρούνται τα παρακάτω:

- Γεωμετρικά αντικείμενα ραβδογράμματος: Στο συγκεκριμένο γράφημα τα αντικείμενα παίρνουν τη μορφή ράβδων. Κάθε ράβδος συμβολίζει ένα σύνολο τιμών. Πιο συγκεκριμένα, τον Μ.Ο. όλων των τιμών από τις απαντήσεις των φοιτητών για κάθε ερώτηση. Κάθε ζεύγος ράβδων διαφορετικού χρώματος, αναλογεί σε μία ερώτηση. Η πρώτη ράβδος του ζεύγους αποτυπώνει τις τιμές απαντήσεων για το σύνολο των μαθημάτων του Προγράμματος Σπουδών που ανήκει το μάθημα το οποίο με τη σειρά του αποτυπώνεται στη δεύτερη ράβδο του ζεύγους.
- Πλαίσιο (frame): Στο γράφημα φαίνεται πως στον άξονα x αντιστοιχίζονται όλες οι ερωτήσεις με την κωδικοποιημένη τους ονομασία. Στον άξονα y φαίνεται η κλίμακα βαθμολόγησης για την αξιολόγηση των μαθημάτων (μηδέν έως πέντε).
- Κλίμακες (scales): Σε κάθε ζευγάρι ράβδων υπάρχει μία ράβδος με πορτοκαλί χρώμα και μία ράβδος με κυανό χρώμα. Αυτό βοηθάει τον θεατή να καταλάβει την πληροφορία πιο εύκολα. Η ράβδος με χρώμα πορτοκαλί αντιστοιχεί στον Μ.Ο. των απαντήσεων των ερωτηματολογίων όλων των μαθημάτων, και αντίστοιχα η ράβδος με χρώμα κυανό στον Μ.Ο. των απαντήσεων σε ένα συγκεκριμένο μάθημα.
- Τίτλοι: Στο ραβδόγραμμα ξεχωρίζει ο γενικός τίτλος του γραφήματος που αναφέρονται όλα τα στοιχεία αυτού. Αναγράφονται οι τίτλοι των αξόνων x,y οι οποίοι είναι «ΕΡΩΤΗΣΗ» και «ΒΑΘΜΟΛΟΓΙΑ». Τέλος, ξεχωρίζουν οι τίτλοι που αντιστοιχούν σε κάθε χρώμα, ο τίτλος «Όλα (843)» και ο τίτλος «Μάθημα (32)» - ο αριθμός 843 αντιστοιχεί στον αριθμό των συνολικών αξιολογήσεων που συλλέχθηκαν από όλα τα μαθήματα του συγκεκριμένου ΠΣ, και ο αριθμός 32 αντιστοιχεί στον αριθμό των συνολικών αξιολογήσεων που συλλέχθηκαν για το συγκεκριμένο μάθημα.
- Πλήθη απαντήσεων: Στην κορυφή κάθε ράβδου αναγράφεται το πλήθος των φοιτητών που απάντησαν σε κάθε ερώτηση για το ένα μάθημα και αντίστοιχα για όλα τα μαθήματα του ΠΣ.

Από το συγκεκριμένο γράφημα φαίνεται ότι για το μάθημα «1626-1741 Εισαγωγή στην Αναλυτική των Δεδομένων», οι Μ.Ο. των απαντήσεων του ερωτηματολογίου αξιολόγησης του μαθήματος είναι πάνω από τον γενικό Μ.Ο. των απαντήσεων.

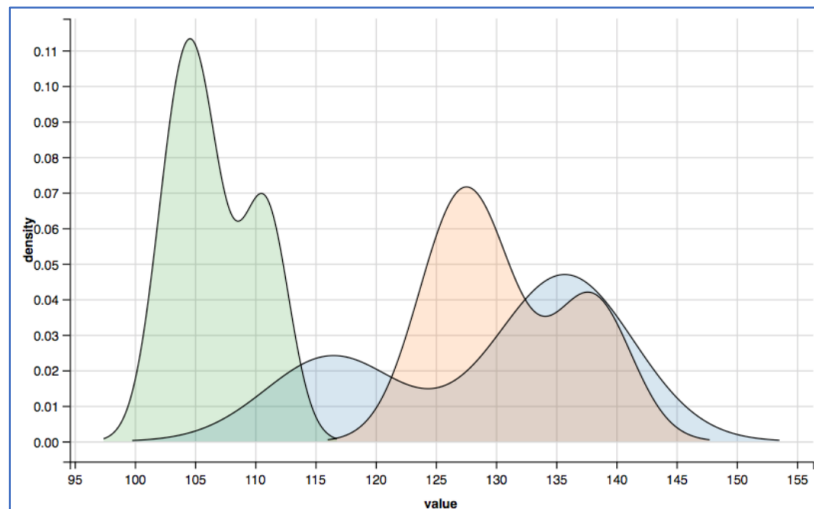
5.4 Γράφημα Πυκνότητας του Δείκτη Ποιότητας

Τα γραφήματα πυκνότητας (Εικόνα 5.6) απεικονίζουν τη συχνότητα και την κατανομή των ποσοτικών τιμών δεδομένων. Ένα διάγραμμα πυκνότητας μπορεί να θεωρηθεί ένα εξομαλυνόμενο ιστόγραμμα καθώς το ιστόγραμμα συγκρίνει ποσότητες χρησιμοποιώντας διακριτές ποσοτικές τιμές ώστε να δημιουργηθούν ομαδοποιήσεις ενώ το διάγραμμα πυκνότητας εξομαλύνει τον θόρυβο επιτρέποντας ομαλότερες κατανομές [B8].



Εικόνα 5.6 Γράφημα πυκνότητας

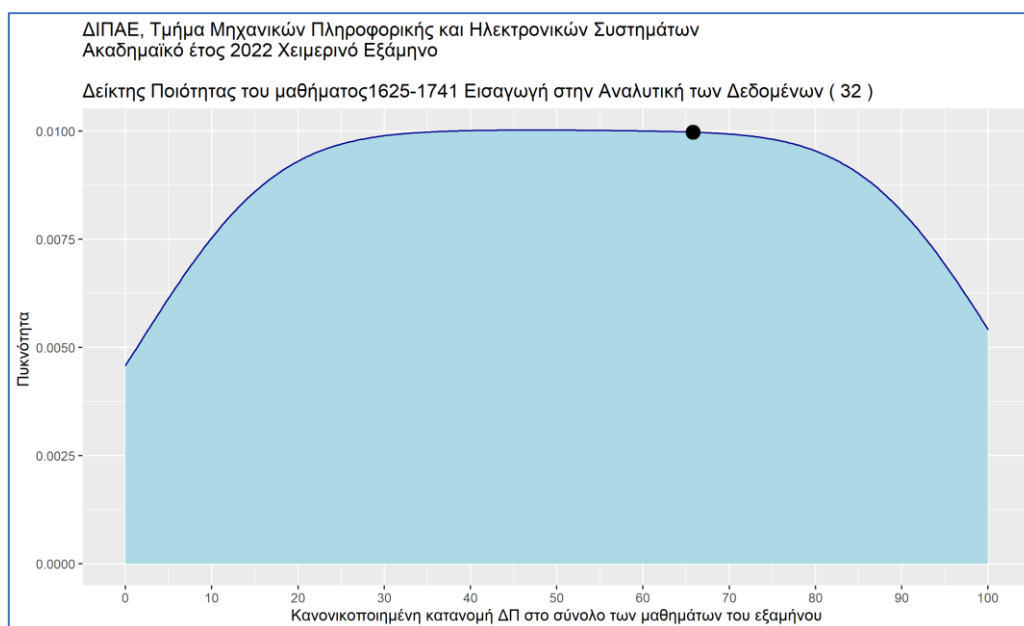
Το γράφημα πυκνότητας τυπικά σχηματίζεται από μια γραμμή, η οποία συνδέει μετρήσεις συχνότητας κάποιας τιμής σε μία κλίμακα. Η γραμμή ανεβαίνει στην κλίμακα όσο αυξάνεται η συχνότητα της τιμής, και μέσω συγκεκριμένων στατιστικών τεχνικών η γραμμή εξομαλύνεται ώστε να σχηματισθεί μία γραμμή χωρίς ακμές. Στη συνέχεια, αν ο ερευνητής το επιθυμεί, η περιοχή εσωτερικά της καμπύλης γεμίζει με χρώμα το οποίο διευκολύνει στην ανάγνωση του γραφήματος. Συχνά, η γραφική παράσταση πυκνότητας περιλαμβάνει διαφορετικές σειρές για διαχωρισμό των στοιχείων σε διακριτές ομαδοποιήσεις κατηγοριών όπως φαίνεται στην Εικόνα 5.7 [B8] [B9].



Εικόνα 5.7 Γράφημα πυκνότητας πολλαπλών σειρών

Οι καμπύλες πυκνότητας είναι ιδιαίτερα χρήσιμες όταν θέλουμε να μελετήσουμε τη σχέση μεταξύ μιας αριθμητικής και μιας ποιοτικής μεταβλητής, δεδομένου ότι δεν καταλαμβάνουν μεγάλη επιφάνεια σ' ένα γράφημα [B10].

Στη παρούσα εργασία το γράφημα πυκνότητας χρησιμοποιήθηκε ως μία επιπλέον τεχνική οπτικοποίησης δεδομένων (Εικόνα 5.8) στο οποίο αποτυπώνεται ο δείκτης ποιότητας ενός μαθήματος σε σχέση με το σύνολο των μαθημάτων του ΠΣ κατά το ακαδημαϊκό έτος που προσφέρθηκε το μάθημα.



Εικόνα 5.8 Γράφημα πυκνότητας κανονικοποιημένης κατανομής του ΔΠ μαθήματος

Για να γίνει καλύτερα κατανοητό το γράφημα πυκνότητας κανονικοποιημένης κατανομής ΔΠ (Εικόνα 5.8) εξηγούνται τα παρακάτω:

- Γεωμετρικά αντικείμενα του γραφήματος: Στο συγκεκριμένο γράφημα τα γεωμετρικά αντικείμενα που αποτυπώνονται είναι η καμπύλη γραμμή του γραφήματος και ένα σημείο. Η καμπύλη αντιπροσωπεύει όλες τις περιπτώσεις των τιμών του ΔΠ των μαθημάτων ενός προγράμματος σπουδών σε ένα εξάμηνο ενώ το σημείο δείχνει τη θέση του μαθήματος πάνω στην κανονικοποιημένη κατανομή των τιμών του ΔΠ.
- Καμπυλότητα: Το ύψος της καμπύλης αντιπροσωπεύει την πυκνότητα των τιμών του ΔΠ των μαθημάτων. Όταν η κλίση της καμπύλης αυξάνεται σημαίνει ότι πλήθος τιμών συνωστίζονται στην περιοχή ενώ όταν το ύψος της καμπύλης μειώνεται σημειώνεται μείωση του πλήθους των τιμών ΔΠ. Στη περίπτωση του παραπάνω γραφήματος αυξημένη συχνότητα συναντάται από τις τιμές ΔΠ τριάντα (30) με (70). Αυτό σημαίνει ότι τα περισσότερα μαθήματα του συγκεκριμένου προγράμματος σπουδών έχουν δείκτη ποιότητας ανάμεσα σε αυτές τις τιμές..
- Τίτλοι: Στο γράφημα φαίνεται ο κυρίως τίτλος του, και οι τίτλοι των αξόνων x,y οι οποίοι είναι «Κανονικοποιημένη κατανομή ΔΠ στο σύνολο των μαθημάτων του εξαμήνου» και «Πυκνότητα».
- Χρώμα: Στο συγκεκριμένο γράφημα το χρώμα δεν αποτυπώνει κάποια πληροφορία αλλά βοηθάει τον αναγνώστη στην καλύτερη γενικότερη κατανόηση.

Από το συγκεκριμένο γράφημα φαίνεται ότι για το μάθημα «1626-1741 Εισαγωγή στην Αναλυτική των Δεδομένων» ο ΔΠ του μαθήματος είναι περίπου εξήντα πέντε (65). Όπως φαίνεται πολλά ακόμη μαθήματα έχουν παρόμοιο ή και ίδιο ΔΠ καθώς στον άξονα y, της πυκνότητας, το σημείο που αντιπροσωπεύει το μάθημα παίρνει την πιο υψηλή τιμή πυκνότητας.

Τα δύο γραφήματα των Εικόνων 5.5 και 5.8 παράχθηκαν χρησιμοποιώντας την υπηρεσία IEE_StAC. Η διαδικασία εκτύπωσης θα περιγραφεί στο Κεφάλαιο 6.

5.4.1 Λίγα Λόγια για τον Δείκτη Ποιότητας

Ο δείκτης ποιότητας είναι ένα εργαλείο διάγνωσης που έχει ως στόχο την παρουσίαση της ποιότητας ενός μαθήματος ενός προγράμματος σπουδών σε σχέση με τα υπόλοιπα μαθήματα του ίδιου ΠΣ.

Ο υπολογισμός του ΔΠ γίνεται μέσω της αντιπαραβολής του Μ.Ο. των τιμών των απαντήσεων από συγκεκριμένες ερωτήσεις της αξιολόγησης κάποιου μαθήματος και του Μ.Ο. των τιμών για όλα τα μαθήματα. Από το σύνολο των ερωτήσεων του ερωτηματολογίου που αναφέρθηκαν στο Κεφάλαιο 2, επιλέχθηκαν οι ακόλουθες οκτώ (8) από τις οποίες υπολογίστηκε ο Μ.Ο.:

- q11: Οι στόχοι του μαθήματος είναι σαφείς.
- q12: Η ύλη που καλύφθηκε ανταποκρίνεται στους στόχους του μαθήματος.
- q13: Η ύλη είναι καλά οργανωμένη.
- q14: Το εκπαιδευτικό υλικό (κύριο βιβλίο, διαφάνειες, σημειώσεις κλπ) βοηθάει στην κατανόηση του μαθήματος.
- q15: Στο μάθημα γίνεται αναφορά στην ύλη άλλων μαθημάτων.
- q16: Στο μάθημα γίνεται σύνδεση γνώσεων με άλλα μαθήματα.
- q17: Το επίπεδο δυσκολίας του μαθήματος είναι ανάλογο του έτους του.
- q18: Ο αριθμός των Διδακτικών Μονάδων του μαθήματος είναι ανάλογες με το φόρτο εργασίας.

Οι συγκεκριμένες ερωτήσεις αφορούν αποκλειστικά το περιεχόμενο του μαθήματος και θεωρείται ότι οι απαντήσεις τους κρίνουν τον βαθμό ποιότητας του μαθήματος. Όσο πιο υψηλή βαθμολογία έχουν τόσο πιο ποιοτικό θεωρείται το μάθημα.

Για να γίνει καλύτερα κατανοητός ο τρόπος υπολογισμού του ΔΠ αρκεί να εξηγήσουμε τον τύπο Rainbow Ranking της Εικόνας 5.9. Το ΔΠ_c συμβολίζει τον δείκτη ποιότητας για ένα μάθημα. Ο υπολογισμός του γίνεται ως εξής [AP2]:

- Έστω ότι θέλουμε να βρούμε τον ΔΠ ενός μαθήματος ΔΠ_c, σε σχέση με το πλήθος των μαθημάτων C.
- Το C είναι το πλήθος των μαθημάτων της λίστας κατάταξης μαθημάτων που ανήκουν στο ίδιο ΠΣ.
- Η λίστα κατάταξης μαθημάτων είναι μία λίστα ταξινομημένη κατά φθίνουσα τάξη, βάση του Μ.Ο των τιμών των απαντήσεων του ερωτηματολογίου αξιολόγησης για τις ερωτήσεις q1 έως και q8.
- Το N_{above(c)} είναι το πλήθος των μαθημάτων που βρίσκονται πάνω από το επιλεγόμενο Μάθημα στη δεδομένη λίστα κατάταξης, άρα με καλύτερες τιμές Μ.Ο..
- Το N_{tie(c)} είναι το πλήθος των μαθημάτων που έχουν την ίδια τιμή Μ.Ο. με το επιλεγόμενο μάθημα.
- Ο τύπος της Εικόνας 5.9 υπολογίζει την κανονικοποιημένη τιμή του δείκτη ποιότητας του μαθήματος (ΔΠ_c) στο διάστημα (0, 100].

$$\Delta\Pi_c = 100 - 100 \left(\frac{|N_{above(c)}|}{|C|} + \frac{|N_{tie(c)}|}{2|C|} \right)$$

Εικόνα 5.9 Τύπος Rainbow Ranking

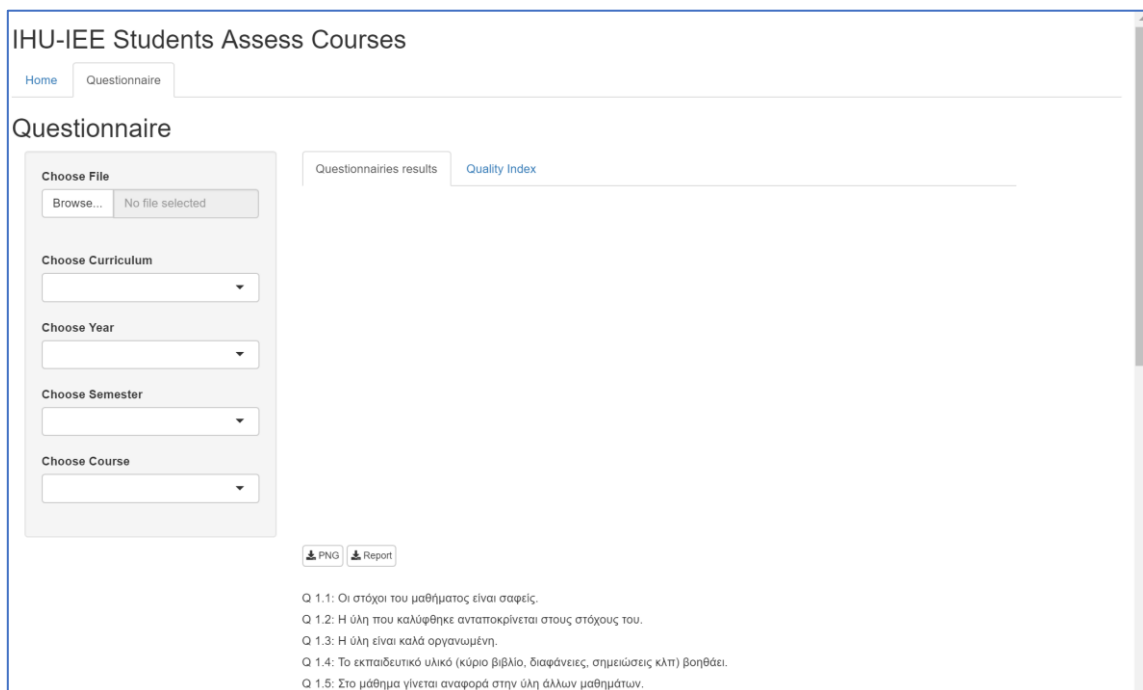
Κεφάλαιο 6ο: Η Υπηρεσία IEE_StAC

Όπως ήδη αναφέρθηκε, η εν λόγω πτυχιακή εργασία επιχειρεί να δημιουργήσει μια διαδικτυακή εφαρμογή, φιλική προς τον χρήστη, που να του επιτρέπει με προσιτό τρόπο να οπτικοποιεί δεδομένα που προέρχονται από την αξιολόγηση των μαθημάτων από τους φοιτητές.

Σε αυτό το κεφάλαιο, εφόσον έχει παρουσιαστεί αναλυτικά η διαδικασία επεξεργασίας πρωτογενών δεδομένων, ο σχεδιασμός της βάσης δεδομένων, και η δημιουργία αλγορίθμων σχεδίασης γραφημάτων οπτικοποίησης της πληροφορίας, παρουσιάζεται το εργαλείο που θα επιτρέψει αυτή την οπτικοποίηση να συμβεί. Το εργαλείο αυτό είναι η διαδικτυακή υπηρεσία IEE Students Assess Courses.

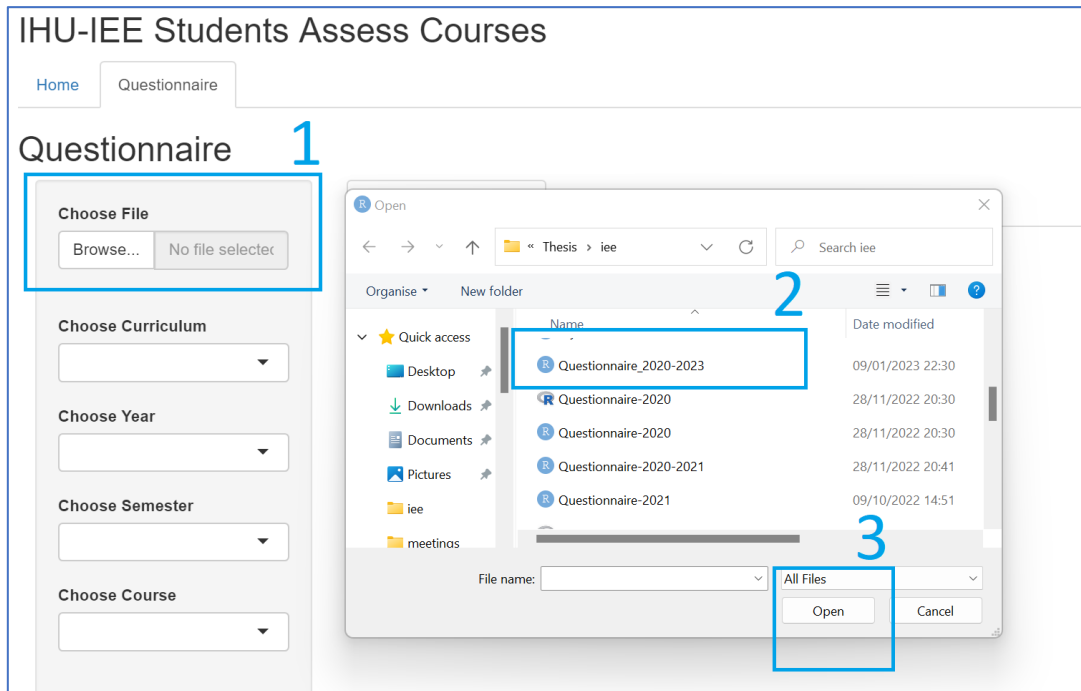
6.1 Με μια Ματιά

Η υπηρεσία που δημιουργήθηκε στο πλαίσιο της πτυχιακής εργασίας, υλοποιήθηκε αξιοποιώντας το εργαλείο της γλώσσας R, την βιβλιοθήκη R Shiny. Αυτή η τεχνολογία καλύπτει κάθε ανάγκη για οπτικοποίηση και διάδραση ανθρώπου μηχανής, ενώ ταυτόχρονα επιτρέπει στο λογισμικό να προσφέρεται διαδικτυακά χωρίς κόστος.



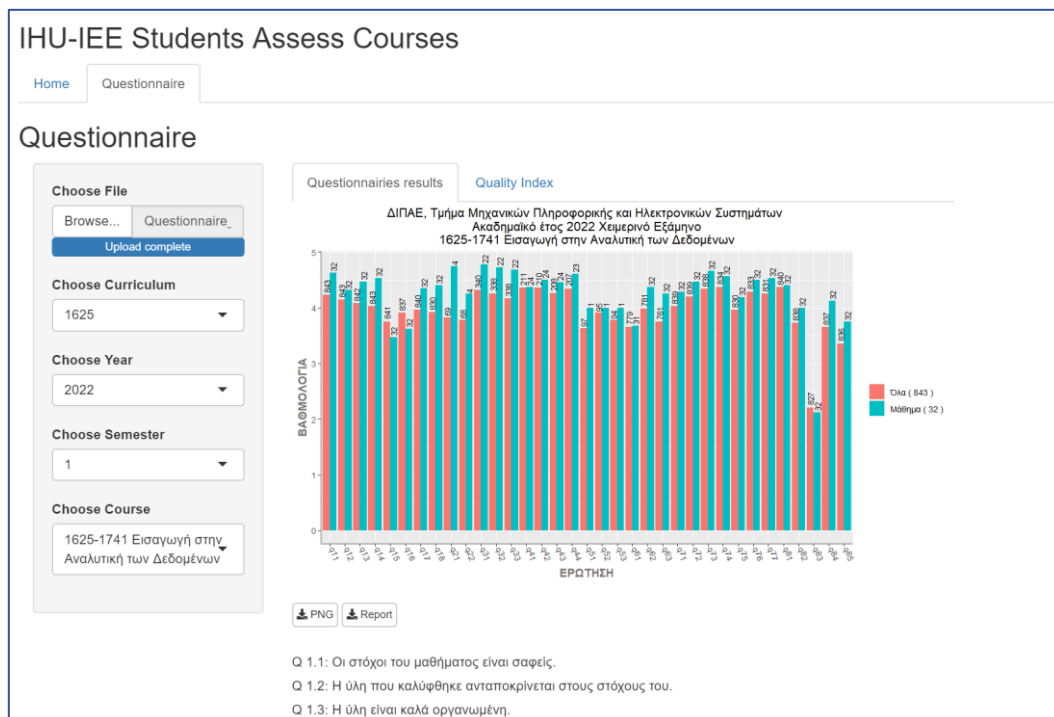
Εικόνα 6.1 Η σελίδα Questionnaire της υπηρεσίας IEE_StAC.

Όταν κάποιος χρησιμοποιεί την υπηρεσία βλέπει ότι αυτή αποτελείται από μία σελίδα με δύο βασικές καρτέλες «Home» και «Questionnaire», όπως φαίνεται στην Εικόνα 6.1. Η καρτέλα «Questionnaire» είναι το περιβάλλον στο οποίο παρουσιάζεται η οπτικοποίηση των δεδομένων μέσω γραφημάτων. Στο αριστερό κομμάτι της καρτέλας υπάρχει το διαδραστικό μενού της εφαρμογής στο οποίο ο χρήστης μπορεί να φορτώσει ένα αρχείο, με την χρήση του κουμπιού «Browse...» (Εικόνα 6.2) και να επιλέξει τιμές πεδίων με διαφορετικές μεταβλητές, εφόσον έχει φορτωθεί το αρχείο δεδομένων. Στο κάτω μέρος της σελίδας ο χρήστης μπορεί να δει την ολοκληρωμένη λίστα των ερωτήσεων του ερωτηματολογίου αξιολόγησης, προς διευκόλυνση του.



Εικόνα 6.2 Φόρτωση ΒΔ στην υπηρεσία IEE_StAC.

Στο δεύτερο επίπεδο της καρτέλας «Questionnaire» έχουν σχεδιαστεί δύο ακόμη καρτέλες, η καρτέλα «Questionnaires results» και οι καρτέλα «Quality index». Η πρώτη καρτέλα αφορά το ραβδόγραμμα αξιολογήσεων ενώ η δεύτερη καρτέλα το γράφημα πυκνότητας του ΔΠ. Ο χρήστης επιλέγει την καρτέλα που επιθυμεί, ανάλογα με την πληροφορία που χρειάζεται. Στην Εικόνα 6.3 παρουσιάζεται η σελίδα που εμφανίζεται εφόσον ο χρήστης έχει φορτώσει τη βάση δεδομένων (Εικόνα 6.2) και έχει επιλέξει τις μεταβλητές που επιθυμεί.



Εικόνα 6.3 Η σελίδα Questionnaire της υπηρεσίας IEE_StAC αφότου φορτωθεί η ΒΔ και επιλεγούν οι μεταβλητές.

Στο επόμενο υποκεφάλαιο γίνεται αναλυτική περιγραφή των επιμέρους λειτουργιών της υπηρεσίας IEE_StAC.

6.2 Λειτουργίες της Υπηρεσίας

Εφόσον η υπηρεσία IEE_StAC έχει δημιουργηθεί για να υποστηρίξει το έργο της ΟΜ.Ε.Α. του Τμήματος Μηχανικών Πληροφορικής και Ηλεκτρονικών Συστημάτων (ΜΠΗΣ) της Σχολής Μηχανικών του Διεθνούς Πανεπιστημίου της Ελλάδος (ΔΙΠΑΕ), το περιεχόμενο της σελίδας αφορά αποκλειστικά το ακαδημαϊκό προσωπικό του Τμήματος. Αυτή η πληροφορία καθορίζει το επίπεδο πρόσβασης στην εφαρμογή από χρήστες.

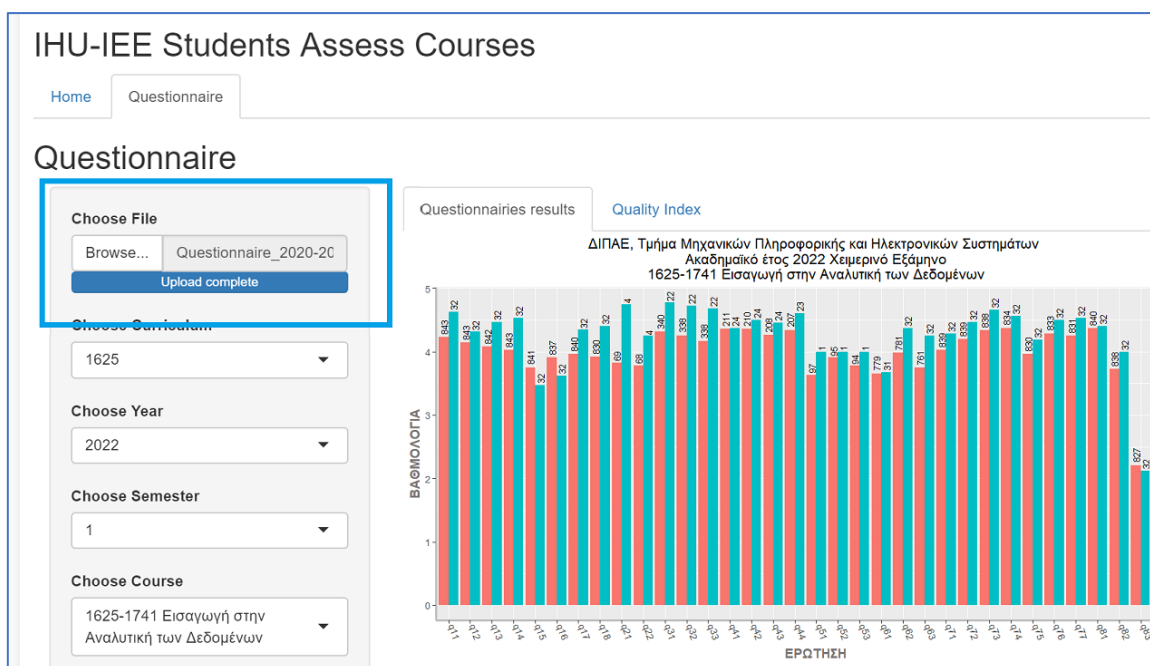
6.2.1 Σύνδεση στην Υπηρεσία

Η υπηρεσία είναι προσβάσιμη μέσω της ιστοσελίδας [IEE Students Assess Courses](https://ihu-quase.shinyapps.io/-iee/) (https://ihu-quase.shinyapps.io/-iee/), η οποία χρησιμοποιεί τον διακομιστή του εργαλείου Shinyapps.io.-Ο σύνδεσμος είναι προσωρινός καθώς μελλοντικά η υπηρεσία θα λειτουργεί στον διακομιστή της Σχολής. Κάθε φορά που κάποιο μέλος του ακαδημαϊκού προσωπικού επιθυμεί πρόσβαση, αρκεί να χρησιμοποιήσει τον σύνδεσμο. Αυτή η απλότητα στην πρόσβαση δίνει την ευκολία στον καθένα που γνωρίζει τον σύνδεσμο να επισκέπτεται τη σελίδα.

6.2.2 Φόρτωση Βάσης Δεδομένων

Η σύνδεση στη σελίδα IEE_StAC είναι μεν απαραίτητη για τη χρήση του λογισμικού αλλά δεν αρκεί. Απαραίτητο στοιχείο για την αξιοποίηση της εφαρμογής είναι η βάση δεδομένων που έχει δημιουργηθεί με σκοπό την χρήση της από την υπηρεσία.

Εφόσον ο επισκέπτης της ιστοσελίδας αποκτήσει πρόσβαση στη βάση δεδομένων, η οποία θα πρέπει να σταλεί σε αυτόν προσωπικά, τότε, χρησιμοποιώντας το κουμπί «Browse...» μπορεί να την φορτώσει στο λογισμικό (Εικόνα 6.3).



Εικόνα 6.3 Σημείο σελίδας όπου υπάρχει η λειτουργία φόρτωσης αρχείου

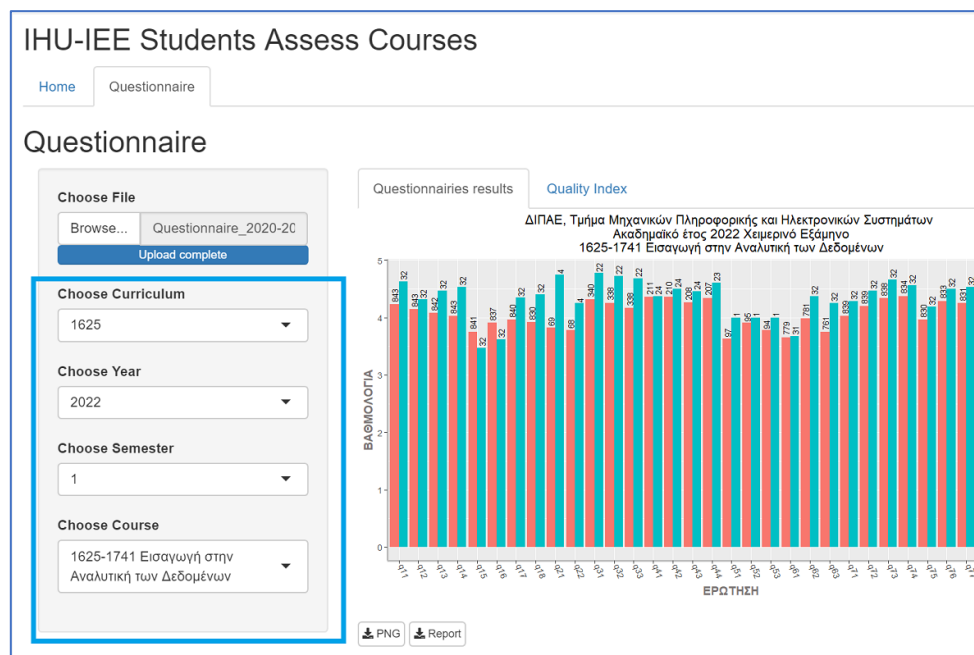
Κατά τον αρχικό σχεδιασμό της υπηρεσίας, η ΒΔ ήταν μέρος του λογισμικού και δεν υπήρχε ανάγκη φόρτωσης της από τον επισκέπτη. Για λόγους ασφαλείας όμως, αποφασίστηκε ότι εφόσον η οπτικοποιημένη πληροφορία αφορά αποκλειστικά τους καθηγητές και την ΟΜ.Ε.Α., τότε τα δεδομένα πρέπει να είναι προσβάσιμα μόνο από τους ίδιους. Έτσι, προστέθηκε ένα αδιάβλητο επίπεδο ασφαλείας αφού ο εκάστοτε καθηγητής οφείλει να φορτώνει ο ίδιος την ΒΔ που θα λαμβάνει από τον υπεύθυνο αξιολογήσεων. Αυτό σημαίνει ότι η υπηρεσία, αν και είναι διαθέσιμη διαδικτυακά, είναι λειτουργική μόνο για τους κατόχους της σχετικής βάσης δεδομένων.

6.2.3 Επιλογή Μεταβλητών

Ο χρήστης, εφόσον φορτώσει την βάση δεδομένων στη διαδικτυακή υπηρεσία, θα διαπιστώσει ότι τα πεδία μεταβλητών αποκτούν τιμές. Τα πεδία μεταβλητών είναι τέσσερα, όπως φαίνεται στην Εικόνα 6.4. Όλα τα στοιχεία επιλογής που εμφανίζονται είναι στοιχεία της ΒΔ τα οποία με την βοήθεια της γλώσσας R έχουν περάσει στο περιβάλλον της εφαρμογής.

Στο πρώτο πεδίο, «Choose Curriculum», ο χρήστης επιλέγει το Πρόγραμμα Σπουδών που τον ενδιαφέρει. Τα Προγράμματα συμβολίζονται με τους κωδικούς τους. Το δεύτερο πεδίο, το «Choose Year» αφορά την ακαδημαϊκή χρονιά κατά την οποία πραγματοποιήθηκε η διδασκαλία του μαθήματος. Οι διαθέσιμες χρονιές είναι τρεις, από το 2020 έως και το 2022. Στο επόμενο πεδίο, το «Choose Semester», παίρνει τις τιμές «1» και «2» για το χειμερινό και εαρινό εξάμηνο αντίστοιχα. Εφόσον ο χρήστης επιλέξει τις επιθυμητές μεταβλητές στα τρία παραπάνω πεδία, τότε εμφανίζονται αυτόματα οι δυνατές επιλογές του τελευταίου πεδίου «Choose Course». Η επιλογή στο πεδίο αυτό αφορά τα μαθήματα που αναλογούν στον συνδυασμό επιλογών του πρώτου, δεύτερου και τρίτου πεδίου. Αν οι επιλογές των τριών πεδίων αλλάξουν τότε αλλάζουν δυναμικά και οι επιλογές του τέταρτου πεδίου. Ο χρήστης μπορεί να επιλέξει το μάθημα που τον ενδιαφέρει μέσα από την ολοκληρωμένη λίστα μαθημάτων.

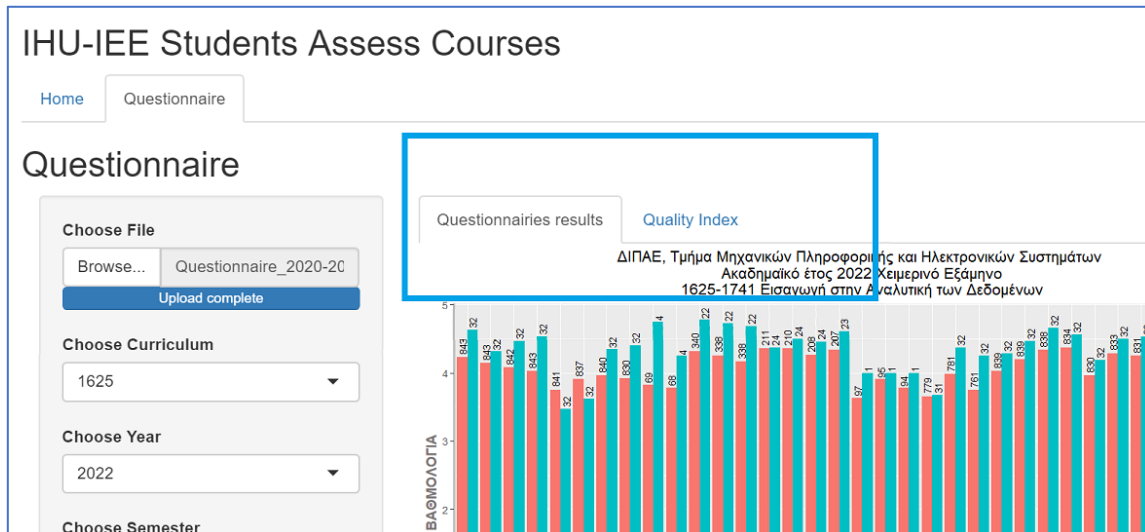
Όταν έχουν επιλεγεί και τα τέσσερα πεδία, εμφανίζεται στο παράθυρο το γράφημα που αφορά το μάθημα της επιλογής του χρήστη.



Εικόνα 6.4 Σημείο σελίδας όπου υπάρχει η λειτουργία επιλογής μεταβλητών

6.2.4 Επιλογή και Εμφάνιση των Γραφημάτων

Ο χρήστης έχει δύο επιλογές όσον αφορά τους τύπους γραφημάτων που μπορεί να παράξει μέσω της υπηρεσίας IEE_StAC. Οι δύο επιλογές είναι είτε το ραβδόγραμμα αξιολογήσεων είτε το γράφημα πυκνότητας του δείκτη ποιότητας. Η πρώτη καρτέλα ονομάζεται «Questionnaires results» και σε αυτήν ο χρήστης μπορεί να εμφανίσει τα ραβδογράμματα αξιολογήσεων για όσα μαθήματα επιθυμεί. Η δεύτερη καρτέλα, η «Quality index», αφορά το γράφημα πυκνότητας του δείκτη ποιότητας του μαθήματος σε σύγκριση με όλα τα μαθήματα του ίδιου ΠΣ.

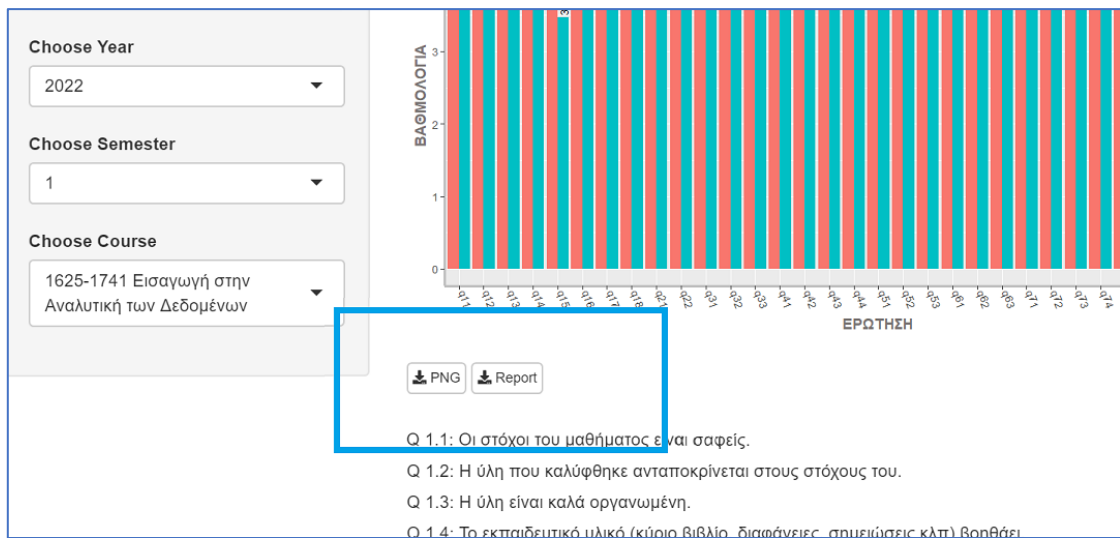


Εικόνα 6.5 Σημείο σελίδας όπου υπάρχει η λειτουργία επιλογής γραφήματος

Εφόσον ο χρήστης φορτώσει την ΒΔ, το λογισμικό την μετατρέπει σε πλαίσιο δεδομένων (data frame) και περνάει τιμές στο δυναμικό μενού. Ύστερα, ο χρήστης επιλέγει τις μεταβλητές που επιθυμεί, και το λογισμικό φιλτράρει τα στοιχεία της βάσης, σύμφωνα με τις μεταβλητές που έχουν επιλεγεί. Ως αποτέλεσμα, η ΒΔ συγκεντρώνει μόνο τα στοιχεία του μαθήματος που επέλεξε ο χρήστης. Στη συνέχεια, καλείται ο αλγόριθμος που αξιοποιεί τα στοιχεία της ανανεωμένης ΒΔ και σχηματίζει το οπτικοποιημένο αποτέλεσμα.

6.2.5 Εκτύπωση των Γραφημάτων και της Αναφοράς του Μαθήματος

Η υπηρεσία εκτός της οπτικοποίησης των δεδομένων δίνει τη δυνατότητα εκτύπωσης των γραφημάτων που ενδιαφέρουν τον χρήστη σε μορφή εικόνας (αρχείο .png) μέσω του κουμπιού «PNG» όπως φαίνεται στην Εικόνα 6.6. Εκτός της εκτύπωσης του αρχείου εικόνας, το αποτέλεσμα της οποίας είναι η Εικόνα 5.5 και η Εικόνα 5.8, ο επισκέπτης της σελίδας μπορεί να εκτυπώσει μία αναφορά του μαθήματος σε τύπο αρχείου MS Word με τη βοήθεια της τεχνολογίας R Markdown. Στην αναφορά υπάρχει ένα περιγραφικό κείμενο που σχετίζεται με την διαδικασία αξιολόγησης του μαθήματος καθώς και το γράφημα αξιολόγησης του μαθήματος.



Εικόνα 6.6 Σημείο σελίδας όπου υπάρχει η λειτουργία λήψης διαγραμμάτων σε μορφή εικόνας ή αναφοράς

Κεφάλαιο 7ο: Προβλήματα που Αντιμετωπίστηκαν και το Μέλλον της Υπηρεσίας

Κατά τη διάρκεια εκπόνησης της πτυχιακής εργασίας εμφανίστηκαν αρκετά προβλήματα τα οποία αντιμετωπίστηκαν βάσει των εργαλείων και γνώσεων που υπήρχαν εκείνη την στιγμή. Σε κάθε έργο υπάρχουν περιθώρια βελτίωσης και η συγκεκριμένη εργασία δεν αποτελεί εξαίρεση. Όσο τα εργαλεία και οι πηγές πληροφόρησης για τις τεχνολογίες που χρησιμοποιήθηκαν πληθαίνουν και βελτιώνονται τόσο θα ανοίγουν νέοι δρόμοι για νέες βελτιωμένες εκδόσεις της υπηρεσίας. Σε αυτό το κεφάλαιο παρουσιάζονται κάποια από τα προβλήματα που αντιμετωπίστηκαν αλλά και αυτά που επιδέχονται βελτίωσης.

7.1 Προβλήματα και Λύσεις

Από την αρχή του χρονικού εκπόνησης της εργασίας, τα διλήμματα ήταν πολλά για την υλοποίηση μιας διαδικασίας με ένα τρόπο, έναντι κάποιου άλλου. Σ' αυτή την παράγραφο αναλύονται μόνο κάποιες περιπτώσεις που θεωρήθηκαν σημαντικές και δεκτικές βελτίωσης για ένα καλύτερο αποτέλεσμα.

7.1.1 Τίτλοι και Στοιχεία Μαθημάτων των Δεδομένων

Συχνά, τα αρχεία δεδομένων των αξιολογήσεων που λαμβάνονταν από τη ΜΟ.Δ.Ι.Π. ανά εξάμηνο, μετά το πέρας της περιόδου αξιολόγησης, περιείχαν λάθη που οδηγούσαν σε προβλήματα όπως έχει εξηγηθεί στο Κεφάλαιο 2. Η υφιστάμενη λύση αυτού του προβλήματος ήταν ο έλεγχος των στοιχείων από άνθρωπο και η διόρθωση λαθών «με το χέρι». Η λύση αυτή δεν είναι ιδανική καθώς είναι χρονοβόρα, και επιπλέον δεν εγγυάται ότι το τελικό αποτέλεσμα είναι αλάνθαστο.

Η κωδικοποίηση και τυποποίηση που αφορούν στα μαθήματα και τους διδάσκοντες θα πρέπει να υποστηρίζονται από το πληροφοριακό σύστημα της ΜΟ.Δ.Ι.Π. Σήμερα, ο διδάσκων, το κάθε ένα μάθημα, καλείται να δηλώσει ο ίδιος τον κωδικό και τον τίτλο του μαθήματος. Αναπόφευκτα συμβαίνουν λάθη τα οποία στη συνέχεια δυσκολεύουν τη διαδικασία επεξεργασίας των δεδομένων από την εφαρμογή IEE_StAC. Η ΟΜ.Ε.Α. του τμήματος έχει ήδη επισημάνει αυτό το πρόβλημα στον διαχειριστή του πληροφοριακού συστήματος της ΜΟ.Δ.Ι.Π.. Σε μελλοντικές εκδόσεις του συγκεκριμένου συστήματος οι κωδικοί και οι τίτλοι των μαθημάτων θα επιλέγονται από λειτουργικά μενού της εφαρμογής της ΜΟ.Δ.Ι.Π. και δε θα εισάγονται με το χέρι από τους διδάσκοντες.

7.1.2 Τα Εργαστήρια και η Αξιολόγηση τους

Το εργαστηριακό μέρος ενός μαθήματος αξιολογείται όπως και το θεωρητικό μέρος, και τα αποτελέσματα της αξιολόγησης συλλέγονται και αποθηκεύονται με τον ίδιο τρόπο. Η επεξεργασία αυτών των δεδομένων χρειάζεται μία ειδική διαχείριση καθώς το εργαστηριακό μέρος αποτελείται από πολλά τμήματα και διαφορετικούς καθηγητές διαφοροποιώντας την διαδικασία οπτικοποίησης της πληροφορίας και παραγωγής γραφημάτων σε σχέση με τις θεωρίες. Γι' αυτό το λόγο αποφασίστηκε ότι η συγκεκριμένη επεξεργασία πρέπει να διαχωριστεί από τη διαδικασία επεξεργασίας των θεωριών και μπορεί να γίνει στο μέλλον.

7.1.3 Διάθεση της Βάσης Δεδομένων

Όπως περιγράφηκε στο Κεφάλαιο 5 ο χρήστης της υπηρεσίας για να έχει πρόσβαση στις λειτουργίες αυτής, πρέπει να έχει στη διάθεση του τη βάση δεδομένων ώστε να την φορτώσει στη σελίδα. Αυτή η

διαδικασία προϋποθέτει ότι το ακαδημαϊκό προσωπικό έχει πάντα στα χέρια του την τελευταία έκδοση της ΒΔ για να έχει πρόσβαση σε σύγχρονη πληροφορία.

Σε μία βελτιωμένη έκδοχή της υπηρεσίας ο χρήστης θα έχει πρόσβαση στα δεδομένα και γραφήματα που τον αφορούν κατά την είσοδο του στην ιστοσελίδα ενώ ταυτόχρονα τα δεδομένα θα παραμένουν ασφαλή από κακόβουλους χρήστες στον ιστό. Για να συμβεί αυτό η υπηρεσία θα πρέπει είτε να φιλοξενηθεί σε έναν διακομιστή της σχολής που μόνο άμεσα εμπλεκόμενοι έχουν πρόσβαση ή διαφορετικά να προστεθεί ένα επιπλέον βήμα για την είσοδο κάποιου στην σελίδα, μέσω διαπιστευτηρίων.

7.2 Επόμενα Βήματα

Κατά τη δημιουργία της εφαρμογής έγινε προσπάθεια ο κώδικας που χρησιμοποιήθηκε να είναι καλά δομημένος ώστε να παραμείνει κατανοητός και ανοιχτός σε αλλαγές και προσθήκες με βάσιμη προσδοκία τη διαρκή βελτίωση της παραχθείσας εφαρμογής.

Τα αναφερθέντα προβλήματα και οι προτάσεις για λύσεις τους, θα μπορούσαν να είναι μέρος μία νέας έκδοσης της υπηρεσίας η οποία θα προσφέρει μία καλύτερη εμπειρία χρήστη αλλά και εμπλουτισμένη πληροφορία μέσω σημαντικών προσθηκών. Ακόμη, θα μπορούσαν να παραχθούν επιπλέον γραφικά με διαφορετικές παραμέτρους όπως η σύγκριση αποτελεσμάτων για το σύνολο των μαθημάτων ενός καθηγητή.

Μία επιπλέον μεταβλητή που αξίζει να μελετηθεί είναι τα ακαδημαϊκά έτη. Η βάση δεδομένων φιλοξενεί πλέον δεδομένα από τρία διαφορετικά έτη, για την ακρίβεια δύο έτη και ένα εξάμηνο, αξίζει να σημειωθεί ότι όταν ξεκίνησε η προσπάθεια περάτωσης της πτυχιακής εργασίας υπήρχαν δεδομένα τριών εξαμήνων. Χρησιμοποιώντας τα ακαδημαϊκά έτη ως μεταβλητή για τη δημιουργία ενός γραφήματος ανά μάθημα, θα μπορούσε ν' αποτυπωθεί μία ίσως πιο αντιπροσωπευτική πληροφορία για την εξέλιξη ενός μαθήματος στον χρόνο.

Ένα σημαντικό επόμενο βήμα για την αξιοποίηση της υπηρεσίας είναι η προσθήκη αυτής ή μέρους της, στο περιβάλλον της ολοκληρωμένης πλατφόρμας πληροφόρησης που αφορά το ακαδημαϊκό προσωπικό του Τμήματος η οποία δημιουργείται από τον υποψήφιο διδάκτορα κ. Κωνσταντίνο Κελεσίδη υπό την επίβλεψη του καθηγητή κ. Δημητρίου Δέρβου. Επιπρόσθετα, η εφαρμογή θα μπορούσε να φιλοξενηθεί στον διακομιστή του Τμήματος.

Επιπλέον, αυτή η υπηρεσία, μετά από κάποιες αλλαγές, θα μπορούσε να αξιοποιηθεί και από άλλα Τμήματα του Πανεπιστημίου. Απαραίτητη βέβαια είναι η χρήση μίας ΒΔ που αφορά τα μαθήματα και τις αξιολογήσεις του ανάλογου Τμήματος η οποία να έχει ίδια μορφή με αυτή που περιγράφεται στην εργασία.

Κεφάλαιο 8ο: Επίλογος

Τα δεδομένα, η επεξεργασία τους και η οπτικοποίηση τους παρουσιάζουν ολοένα και μεγαλύτερη ζήτηση σε όλους τους τομείς. Επιπλέον, εταιρίες, κυβερνήσεις και λοιποί οργανισμοί επενδύουν συνεχώς στην καλύτερη πληροφόρηση που οδηγεί στη λήψη ορθών αποφάσεων. Έτσι, όλο και περισσότεροι ενδιαφέρονται για να αξιοποιήσουν τα δεδομένα τους προκειμένου να διαθέτουν χρήσιμες πληροφορίες.

Το έργο της εξαγωγής μοτίβων από διάφορες πηγές δεδομένων οδηγεί στην δυνατότητα ορθών συμπερασμάτων και την πρόβλεψη τάσεων. Επίσης, συχνά ακούγεται στους χώρους του μάνατζμεντ η έκφραση «ό,τι μετριέται βελτιώνεται». Στο πλαίσιο βελτίωσης του ακαδημαϊκού προγράμματος του Τμήματος ΜΠΗΣ η παρούσα εργασία θέλει να συμβάλει στο έργο της ΟΜ.Ε.Α. Η εφαρμογή βάζει ένα μικρό λιθαράκι ώστε οι αξιολογήσεις του εκάστοτε φοιτητή/τριας για τους καθηγητές/τριες και τα μαθήματα που διδάσκουν να παρουσιάζονται με τρόπο εποπτικό που να βοηθά στην κατανόηση της πληροφορίας.

Η εργασία με όλες τις τεχνικές που προσφέρει θέλει να συνεισφέρει στην πρόοδο των υπηρεσιών που προσφέρει η ΟΜ.Ε.Α.. Πέρα από τη δημιουργία της εφαρμογής η οποία αφορά την οπτικοποίηση της πληροφορίας και την ευκολία διάθεσης της, αποτυπώνεται και η πρότυπη διαδικασία δημιουργίας και ανανέωσης της ΒΔ που χρησιμοποιείται από την εφαρμογή.

Με τα παραπάνω εργαλεία και τις τεχνικές που έχει επιπλέον η ΟΜ.Ε.Α. στα χέρια της, υποστηρίζονται καλύτερα οι λειτουργίες της και το όραμα της. Επιπλέον, τα ίδια εργαλεία μπορούν να συντελέσουν στην πρόοδο των υπηρεσιών αξιολόγησης άλλων τμημάτων δίνοντας κίνητρο για δημιουργία νέων υπηρεσιών που θα συμβάλουν στη γενική ανάπτυξη και προσφορά υψηλού επιπέδου ποιότητας ακαδημαϊκών εμπειριών.

Σε κάθε περίπτωση, η συγγράφουσα ελπίζει ότι πρόσθεσε κάποια έστω και μικρή αξία στην ψηφιακή εργαλειοθήκη του Ιδρύματος, στην κατεύθυνση της βελτίωσης των δυνατοτήτων αξιολόγησης του έργου που επιτελείται. Ταυτόχρονα θεωρεί τον εαυτό της τυχερό που είχε την ευκαιρία εποικοδομητικής συνεργασίας με αξιόλογους επιστήμονες σ' ένα αντικείμενο που παρουσιάζει μεγάλο ενδιαφέρον στο παρόν και το μέλλον.

Κεφάλαιο 9ο: Βιβλιογραφία

Books (Paper or Digital)

- [B1] I. Mahdavi, S. Mohebbi, N. Cho, *Electronic Supply Network Coordination in Intelligent and Dynamic Environments Modeling and Implementation*. Iraj Mahdavi, Namjae Cho, Shima Mohebbi, 2011, Pp 44-45.
- [B2] Shinyapps.io team, *Shinyapps.io user guide*. Latest update: Aug 2020.
- [B3] T. Mailund, *R Data Science Quick Reference A Pocket Guide to APIs, Libraries, and Packages*. California: Apress , 2019, Pp 2-3.
- [B4] Yihui Xie, J. J. Allaire, G. Grolemond, *R Markdown: The Definitive Guide*. UK: Chapman & Hall/CRC.
- [B5] Kun Ren, *Learning R Programming* UK, Birmingham: Packt Publishing Ltd. Oct 2026, pp 204-205.
- [B6] Leo M. Gordon, Daniel B. Garrie, *Cybersecurity & the Courthouse. Safeguarding the Judicial Process*. United States: Wolters Kluwer, 2020, Pp 49-50.
- [B7] C. Satyadevi, *Quantitative Techniques*. S.Chand Limited, 2009, Pp 71-80.
- [B8] Andy Kirk, *Data Visualisation A Handbook for Data Driven Design*. SAGE Publications, 2019, pp 154-155.
- [B9] Robert Kabacoff, *R in Action, Third Edition Data Analysis and Graphics with R and Tidyverse*. Manning, May 2022, pp 154-155.
- [B10] Homer White, *Beginning Computer Science with R*. US: Creative Commons Attribution, 2023
- [B11] Claus O. Wilke, *Fundamentals of Data Visualization A Primer on Making Informative and Compelling Figures* US: O'Reilly Media, Inc, 2019.
- [B12] Keith McNulty, *Handbook of Regression Modeling in People Analytics: With Examples in R, Python and Julia*. Chapman & Hall, 2022.
- [B13] Leland Wilkinson, *The Grammar of Graphics*. United States: Springer New York, 2005.

Internet Sites

- [IS1] Internal Evaluation Group, Department of Information and Electronic Engineering - International Hellenic University “Αξιολόγησης (OM.E.A.)”, 2020. Available: <https://omea.iee.ihu.gr/site/> (Accessed: 18 January 2023).
- [IS2] Διεθνές Πανεπιστήμιο της Ελλάδος, “Επίσημη ιστοσελίδα του Πανεπιστημίου, Quality Assurance Unit Home Page”, Available: <https://www.ihu.gr/modip/en/quality-assurance-unit/> (Accessed: 18 January 2023).
- [IS3] Posit Company, Official website of RStudio IDE, [Online]. Available: <https://posit.co/> (Accessed: 20 December 2022).
- [IS4] R Foundation, “The R Project for Statistical Computing”. Available: <https://www.r-project.org/> (Accessed: 20 December 2022).

- [IS5] R Core Team, “R Language Definition”. Available: <https://cran.r-project.org/doc/manuals/r-devel/R-lang.html> (Accessed: 20 December 2022).
- [IS6] CRAN Repository Maintainers, “CRAN Repository Policy”. Available: <https://cran.r-project.org/> (Accessed: 20 December 2022).
- [IS7] R Markdown file, “How R Markdown Works”. Available: <https://rmarkdown.rstudio.com/lesson-2.html> (Accessed: 20 December 2022).
- [IS8] A Matt Cone project, “The Markdown Guide”. Available: <https://www.markdownguide.org/> (Accessed: 20 December 2022).
- [IS9] Pandoc, “Pandoc User’s Guide, Pandoc’s Markdown”. Available: <https://pandoc.org/MANUAL.html#pandocs-markdown> (Accessed: 10 March 2022).
- [IS10] Manny Gimond, “Reading and Writing Data Files”. Available: https://mgimond.github.io/ES218/Week02b.html#Reading_data_files_into_R (Accessed: 20 December 2022).
- [IS11] A. Kipp, “Getting Started with shinyapps.io”, May 2017. Available: <https://shiny.rstudio.com/> (Accessed: 20 December 2022).
- [IS12] Posit Company, Official website of shinyapps.io. Available: <https://www.shinyapps.io/> (Accessed: 20 December 2022).
- [IS13] University of Cincinnati, UC Business Analytics R Programming Guide. Available: <http://uc-r.github.io/> (Accessed: 18 January 2022).

Academic Papers

- [AP1] K. Kelesidis, “Προτυποποίηση αναλυτικής επεξεργασίας δεδομένων βαθμολογιών”, Διπλωματική Εργασία, ΠΜΣ: Ευφυείς Τεχνολογίες Διαδικτύου, Τμήμα Μηχανικών Πληροφορικής και Ηλεκτρονικών, ΔΙΠΙΑΕ, Σεπ. 2020.
- [AP2] K. Kelesidis, D. Dervos, A Sidiropoulos, Quantifying the Difficulty of Academic Course Modules", In Proceedings of the 24th Pan-Hellenic Conference on Informatics, Athens, Greece, Nov 2020, ACM Digital Library, DOI: 10.1145/3437120.3437317.

Online Articles

- [A1] Directions Magazine, “What is Data Analysis and the Role of Data Analyst?”, May. 2018.
- [A2] Simplilearn, “What is R: Overview, its applications and what is R used for”, Nov. 2022.
- [A3] The Council on Quality and Leadership, “12 Reasons Why Data Is Important”.
- [A4] Craig Stedman “Definition: Data Quality”, Dec. 2022.
- [A5] Towards Data Science, “A Comprehensive Guide to the Grammar of Graphics for Effective Visualization of Multi-dimensional Data”, Sep 2018.

ΠΑΡΑΡΤΗΜΑ Α : Αρχείο Κώδικα R «step1_data_prepare»

Το τμήμα κώδικα που χρησιμοποιείται στο «Βήμα 2: Διαχωρισμός Μαθημάτων και Επεξεργασία» του Κεφαλαίου 3.

```
## Αρχικά πρέπει να οριστούν οι μεταβλητές 'Year', 'S' και 'PS' ανάλογα με το αρχείο δεδομένων
# που θέλουμε να επεξεργαστούμε. Δηλώνουμε τη χρονιά, το εξάμηνο και το ΠΣ.

# Εισαγωγή βιβλιοθηκών και πακέτων
library(readxl)

#Ορίζουμε τοπικές ρυθμίσεις ώστε να διαβάζονται τα ελληνικά κείμενα δεδομένων
Sys.setlocale(category = "LC_ALL", locale = "Greek")

#Ορίζουμε το ακαδημαϊκό έτος στη μεταβλητή 'Year'
#Π.χ. αν το ακαδημαϊκό έτος είναι το 2022-23 -> 2022
Year <- 2022

#Δημιουργούμε νέο directory για το νέο ακαδημαϊκό έτος
mydir = paste0(getwd(),"/results/",Year)
dir.create(mydir)

## Ορίζουμε το εξάμηνο στη μεταβλητή 'S'
# 1. Xeimerino
# 2. Earino
S <- 1
mydir = paste0(getwd(),"/results/",Year,"/",S,"/")
dir.create(mydir)

## Ορίζουμε το Πρόγραμμα σπουδών στη μεταβλητή 'PS' Εδώ το ΠΣ δηλώνεται μέσα στο loop for
# 1. PPS
# 2. PMS-ETD
# 3. PMS-EHS
# Προσοχή! στο look for πρέπει να δηλωθούν σωστά τα paths των φακέλων που βρίσκονται τα αρχεία
for(PS in 1:3){
  cat("PS:",PS,"\n")
  if (PS==1){ # PPS
    Data <- read_xlsx("raw data/22-23_Winter_PPS.xlsx" , col_names =FALSE)
```

```

mydir = paste0(getwd(),"/results/",Year,"/",S,"/pps")
} else if (PS==2){ # PMS-ETD
  Data <- read_xlsx("D:/Thesis/iee/raw data/22-23_Winter_MPS_ETD.xlsx", col_names =FALSE)
  mydir = paste0(getwd(),"/results/",Year,"/",S,"/pms-etd")
} else if (PS==3){ # PMS-EIS
  Data <- read_xlsx("D:/Thesis/iee/raw data/22-23_Winter_MPS_EHS.xlsx", col_names =FALSE)
  mydir = paste0(getwd(),"/results/",Year,"/",S,"/pms-eis")
} else {
  cat("Lathos PS")
}

# Αν θέλω να επεξεργαστώ ένα συγκεκριμένο πρόγραμμα σπουδών μπορώ να τρέξω κατευθείαν
# το παρακάτω κομμάτι κώδικα αλλάζοντας την μεταβλητή 'PS'και το path
#PS <- 3
#Data <- read_xlsx("D:/Thesis/iee/raw data/22-23_Winter_MPS_EHS.xlsx", col_names =FALSE)
#mydir = paste0(getwd(),"/results/",Year,"/",S,"/pms-eis")

# Δημιουργώ το directory αν δεν υπάρχει
dir.create(mydir)

# Δηλώνω τον δείκτη με τον οποίο κλείνει κάθε ερωτηματολόγιο,
# Η ερώτηση q85 και για κάποια ερωτηματολόγια η q65
closure_index <- which(grepl("q65|q85", Data$...1))

start<-1 # Δείκτης έναρξης του μαθήματος

# δημιουργία dataframe για κάθε μάθημα
for (value in closure_index) {
  end <- value # Δείκτης λήξης του μαθήματος
  testdata <- as.data.frame(Data[start:end,]) # δημιουργώ dataframe για το μάθημα

  if (PS==1){
    title <- paste0(testdata[1,2]," -- ",testdata[3,1]) # στο PPS έχει μία γραμμή λιγότερη
  } else {
    title <- paste0(testdata[1,2]," -- ",testdata[4,1]) # στα PMS αναφέρεται και το τμήμα
  }
}

```

```

title <- gsub("/", "", title) # αφαιρώ τον χαρακτήρα /
title <- gsub("*", "", title) # αφαιρώ τον χαρακτήρα *
title <- gsub("Διδάσκων:", "", title) # αφαιρώ τον χαρακτήρα το 'Διδάσκων:'
title <- gsub(", ", "_", title) # Αντικαθιστώ το ',' με '_'

# αποθηκεύω τα νέα αρχεία με ονόματα ανάλογα με το αν είναι θεωρίες ή εργαστήρια
if(testdata[1,3]=="θεωρία") {
  path <- paste0(mydir,"/",title,"--theory.csv")
}
if(testdata[1,3]=="εργαστήριο") {
  path <- paste0(mydir,"/",title,"--lab.csv")
}
write.csv(testdata, path, row.names = FALSE)
start<-end+1
}
}

```


ΠΑΡΑΡΤΗΜΑ Β : Αρχείο Κώδικα R «step2_data_prepare»

Το τμήμα κώδικα που χρησιμοποιείται στο «Βήμα 3: Δημιουργία Ενός Νέου Αρχείου .csv για Κάθε Πρόγραμμα Σπουδών» του Κεφαλαίου 3.

```
## Αρχικά πρέπει να οριστούν οι μεταβλητές 'Year', 'S' και 'PS' ανάλογα με το αρχείο δεδομένων
# Παίρνουμε ως δεδομένο ότι τα δεδομένα μας βρίσκονται σε υποφακέλους του φακέλου 'results'.

# Εισαγωγή βιβλιοθηκών και πακέτων
library(dplyr)
library(plyr)
library(reshape2)
library(tidyr)
library(xlsx)
library(stringr)

#Ορίζουμε το ακαδημαϊκό έτος στη μεταβλητή 'Year'
#Π.χ. αν το ακαδημαϊκό έτος είναι το 2022-23 -> 2022
Year <- 2022

## Ορίζουμε το εξάμηνο στη μεταβλητή 'S'
# 1. Xeimerino
# 2. Earino
S <- 1

## Ορίζουμε το Πρόγραμμα σπουδών στη μεταβλητή 'PS' Εδώ το ΠΣ δηλώνεται μέσα στο loop for
# 1. PPS
# 2. PMS-ETD
# 3. PMS-EHS
# Προσοχή! στο loop for πρέπει να δηλωθούν σωστά τα paths των φακέλων που βρίσκονται τα αρχεία
ή
# να χρησιμοποιηθεί η μεταβλητή dir που βρίσκεται σε σχόλιο
for(kkk in 1:3){
  cat("kkk:",kkk,"\n")
}
```

```

PS <- kkk
dir = paste0(getwd(),"/results/",Year,"/",S,"/")
if (PS==1){
  mydir = paste0(getwd(),"/results/",Year,"/",S,"/pps/")
  prog_spoudon="1625"
  tmima="ΠΠΣ"
} else if (PS==2){
  mydir = paste0(getwd(),"/results/",Year,"/",S,"/pms-etd/")
  prog_spoudon="1269"
  tmima="ΠΜΣ Ε.Τ.Δ."
} else if (PS==3){
  mydir = paste0(getwd(),"/results/",Year,"/",S,"/pms-eis/")
  prog_spoudon="1268"
  tmima="ΠΜΣ Ε.Η.Σ."
} else {
  cat("Lathos PS")
}

```

Κρατάμε στη μνήμη τις τοποθεσίες των θεωριών και εργαστηρίων

```
myTheories = list.files(path=mydir, pattern="*--theory.csv", full.names=TRUE)
```

```
myLabs = list.files(path=mydir, pattern="*--lab.csv", full.names=TRUE)
```

Επεξεργασία θεωριών

```
list_of_theories <- list()
```

```
i<-1
```

```
for (value in myTheories) {
```

```
  course<-read.csv(value, header = T) # διαβάζω το αρχείο
```

```
  lineOne = grep("q11",course[,1])-1 # βρίσκω την γραμμή που αρχίζουν οι ερωτήσεις στη ΒΔ
```

```
  course<-course[-2:-lineOne,] # Κρατάω τον τίτλο και τις ερωτήσεις του μαθήματος
```

```
  course$title <- paste0(course[1,2]," ",course[1,1]," -- ",course[1,3]) # κρατάω τον τίτλο του μαθήματος
```

```
  course <- course[-1,] # κρατάω μόνο τις ερωτήσεις/απαντήσεις, αφαιρώ τον τίτλο
```

```

for(k in 2:(ncol(course)-1)){
  course[,k]<-as.numeric(course[,k]) # κάνω όλα τα πεδία απαντήσεων αριθμητικά
}

course$Mean <- rowMeans(course[,2:(ncol(course)-1)], na.rm=TRUE) # Μ.Ο. του μαθήματος

course$C_Counter <- rowSums(!is.na(course[, 2:(ncol(course)-2)])) # πλήθος απαντήσεων για το
μάθημα

list_of_theories[[i]]<-course
i<-i+1
}

concat_theories <- ldply(list_of_theories, data.frame) # μετατροπή λίστας σε data.frame

## Εργαστήρια
list_of_labs <- list()
i<-1
for (value in myLabs) {
  course<-read.csv(value, header = T) # διαβάσω το αρχείο
  lineOne = grep("q11",course[,1])-1 # βρίσκω την γραμμή που αρχίζουν οι ερωτήσεις στη ΒΔ
  course<-course[-2:-lineOne,] # Κρατάω τον τίτλο και τις ερωτήσεις του μαθήματος
  course$Title <- paste0(course[1,2]," ",course[1,1]," -- ",course[1,3]) # κρατάω τον τίτλο του
μάθηματος
  course <- course[-1,] # κρατάω μόνο τις ερωτήσεις/απαντήσεις, αφαιρώ τον τίτλο

  for(k in 2:(ncol(course)-1)){
    course[,k]<-as.numeric(course[,k]) # κάνω όλα τα πεδία απαντήσεων αριθμητικά
  }

  course$Mean <- rowMeans(course[,2:(ncol(course)-1)], na.rm=TRUE) # Μ.Ο. του μαθήματος

  course$C_Counter <- rowSums(!is.na(course[, 2:(ncol(course)-2)])) # πλήθος απαντήσεων για το
μάθημα

```

```

list_of_labs[[i]]<-course
i<-i+1
}
concat_labs <- ldply(list_of_labs, data.frame) # μετατροπή λίστας σε data.frame

dataset_theories<-data.frame()
plot_data_theories<-list()
if( !dim(concat_theories)[1] == 0){
  # Βρίσκουμε τους Μ.Ο. για το σύνολο των μαθημάτων
  questions_theories<-unique(concat_theories[1]) # Μοναδικές τιμές των ερωτήσεων
  dataset_theories<-list()
  for(i in 1:nrow(questions_theories)){

    question_name<-paste0(questions_theories[i,1]) # βρίσκω την ερώτηση
    data<-subset(concat_theories,concat_theories[1]==question_name) #βρίσκω τις γραμμές με την
    ερώτηση
    temp<-data[2:(length(data)-3)] # Κρατάω τις απαντήσεις
    data$AllMean<-mean(as.matrix(temp), na.rm=TRUE) # Μετατρέπω σε matrix και βρίσκω τον
    Μ.Ο. όλων των μαθημάτων για την ερώτηση
    data$A_Counter<-sum(!is.na(temp)) # το πλήθος των απαντήσεων για την ερώτηση σε όλα τα
    μαθήματα
    dataset_theories[[i]]<-data
  }

  ## επεξεργασία τίτλων και δεδομένων για χρήση στο plot
  for (i in 1:length(dataset_theories)) {
    df <- dataset_theories[[i]]
    df<-data.frame( df[1], df[(ncol(df)-4)], df[(ncol(df)-3)], df[(ncol(df)-2)], df[(ncol(df)-1)],
    df[(ncol(df))])
    colnames(df)<-c("Question", "Title", "Mean", "C_counter", "AllMean", "A_Counter")

    dm1 <- melt(df[,c("Question", "Title", "Mean", "AllMean")], id=c("Question", "Title"))
  }
}

```

```

dm2 <- melt(df[,c("Question", "Title", "C_counter", "A_Counter")], id=c("Question", "Title"))
colnames(dm2) <- c("Question", "Title", "variable2", "value2")
dm <- cbind(dm1, dm2[,3:4])
plot_data_theories[[i]]<-dm
}
##
plot_data_theories <- ldply(plot_data_theories, data.frame)
##
plot_data_theories <- plot_data_theories %>% separate(Title, c("Title", "Category"), " -- ")
plot_data_theories <- plot_data_theories %>% separate(Question, c("QN", "Question"), "\\.")
plot_data_theories$ProgSpoudon <- prog_spoudon
plot_data_theories$Year <- Year
plot_data_theories$Semester <- S

}

```

επανάληψη διαδικασίας για τα εργαστήρια

```

dataset_labs<-data.frame()
plot_data_labs<-list()
if( !dim(concat_labs)[1] == 0){
  # Βρίσκω Μ.Ο.
  questions_labs<-unique(concat_labs[1]) # Μοναδικές τιμές ερωτήσεων
  dataset_labs<-list()
  for(i in 1:nrow(questions_labs)){

    question_name<-paste0(questions_labs[i,1]) # Βρίσκω την ερώτηση
    data<-subset(concat_labs,concat_labs[1]==question_name) #βρίσκω τις γραμμές με την ερώτηση
    temp<-data[2:(length(data)-3)] # Κρατάω τις απαντήσεις
    data$AllMean<-mean(as.matrix(temp), na.rm=TRUE) # Μετατρέπω σε matrix και βρίσκω τον
    Μ.Ο. όλων των μαθημάτων για την ερώτηση
    data$A_Counter<-sum(!is.na(temp)) # το πλήθος των απαντήσεων για την ερώτηση σε όλα τα
    μαθήματα
    dataset_labs[[i]]<-data
  }
}

```

```

}

## επεξεργασία τίτλων και δεδομένων για χρήση στο plot
for (i in 1:length(dataset_labs)) {

  df <- dataset_labs[[i]]

  df<-data.frame( df[1], df[(ncol(df)-4)], df[(ncol(df)-3)], df[(ncol(df)-2)], df[(ncol(df)-1)],
df[(ncol(df))] )
  colnames(df)<-c("Question", "Title", "Mean","C_counter","AllMean","A_Counter")

  dm1 <- melt(df[,c("Question", "Title", "Mean","AllMean")], id=c("Question", "Title"))
  dm2 <- melt(df[,c("Question", "Title", "C_counter","A_Counter")], id=c("Question", "Title"))
  colnames(dm2) <- c("Question", "Title", "variable2", "value2")
  dm <- cbind(dm1, dm2[,3:4])
  plot_data_labs[[i]]<-dm
}
##
plot_data_labs <- ldply(plot_data_labs, data.frame)
##
plot_data_labs <- plot_data_labs %>% separate(Title, c("Title", "Category"), "-- ")
plot_data_labs <- plot_data_labs %>% separate(Question, c("QN", "Question"), "\\.")
plot_data_labs$ProgSpoudon <-prog_spoudon
plot_data_labs$Year <- Year
plot_data_labs$Semester <- S

}

plot_data<-rbind(plot_data_theories,plot_data_labs)

```

```

# Χωρίζουμε το όνομα και το ID του μαθήματος
plot_data[c('id', 'Name')] <- str_split_fixed(plot_data$Title, ' ', 2)

wd<-getwd()

#respath<-paste0(dir,"res-",Year,"-",S,".xlsx")
respath<-paste0(getwd(),"/Questionnaire-",Year,"-",PS,"-",S,".xlsx")
write.xlsx(plot_data, file = respath,
           sheetName = prog_spoudon , col.names = TRUE, append = TRUE, row.names = TRUE)

setwd(wd)
}

#read excels (εδώ το path πρέπει να αλλάζει ανάλογα με το που βρίσκονται τα αρχεία και η εντολή θα
πρέπει να τρέχει για κάθε αρχείο που θέλουμε να διαβάσουμε ώστε να περαστεί σε ένα κοινό αρχείο)
data4 <-read_xlsx("D:/Thesis/iee/results/2022/1/Questionnaire-2022-3-1.xlsx", col_names =TRUE, na
= "" )

#combine dataframes (τα ονόματα των dataframes πρέπει να προσαρμόζονται ανάλογα με τα αρχεία
excel που έχουμε περάσει στη μνήμη)
dataall<-rbind(data4,data5,data6)

# Δημιουργία rdata αρχείου
respath<-paste0(getwd(),"/Questionnaire-Year",".rds")
saveRDS(dataall, respath)

my_data2 <- readRDS(respath)

respath<-paste0(getwd(),"/Questionnaire-",Year,".RData")
save(dataall, file = respath)
my_data <- load("Questionnaire-2020-1.RData")

```