*Ηλεκτρονική Μάθηση*

*Εξόρυξη εκπαιδευτικών δεδομένων και Μαθησιακή Αναλυτική*

**Εισαγωγή**

Η Ψηφιακή Επανάσταση που παρατηρείται τα τελευταία χρόνια ήταν αδύνατο να μην επηρεάσει τον τομέα της Εκπαίδευσης. Η Εκπαίδευση(σε όλες τις μορφές της) αποτελεί μια σημαντική διαδικασία αγωγής και μάθησης για τον άνθρωπο. Δηλαδή, η σύγχρονη εκπαίδευση βρίσκεται σε διαρκή εξέλιξη με την ανάπτυξη της τεχνολογίας. Αξίζει να τονιστεί ότι η εισαγωγή της ηλεκτρονικής μάθησης στην εκπαίδευση έχει επαναπροσδιορίσει τον τρόπο με τον οποίο λειτουργεί τόσο ο μαθητής/φοιτητής όσο και ο διδάσκων.

Παράλληλα, μέρος των τεχνολογικών εξελίξεων που αναφέραμε αποτελεί και η ραγδαία αύξηση της κλίμακας και της ποσότητας των δεδομένων που υπάρχουν σε ηλεκτρονική μορφή(εποχή των Big Data). Μείζον ζήτημα για την επιστημονική κοινότητα αποτελεί η αποθήκευση, αλλά και η επεξεργασία-ανάλυση των δεδομένων(Ray,Saeed,2018). Στο χώρο της εκπαίδευσης συγκεκριμένα οι τεχνικές ανάλυσης δεδομένων αναφέρονται με τους όρους Εξόρυξη Εκπαιδευτικών Δεδομένων(ΕΔ) και Μαθησιακή Αναλυτική(ΜΑ).

Η χρήση εργαλείων ηλεκτρονικής μάθησης(για παράδειγμα eclass, Moodle) όπως και η εξ αποστάσεως εκπαίδευση, που έγινε απαραίτητη στην περίοδο του Covid, οδήγησαν στην αύξηση των e-learning πόρων και στη δημιουργία ισχυρών εκπαιδευτικών αποθετηρίων(Romero,Ventura,2020). Αναμφίβολα, τη μεγαλύτερη πρόκληση αποτελεί η μετατροπή των εκπαιδευτικών δεδομένων σε γνώση και ουσία, με στόχο τη βελτίωση της εκπαιδευτικής διαδικασίας, αλλά και την ατομική βελτίωση μαθητή και καθηγητή.

Όσον αφορά το κεντρικό θέμα της εργασίας μας, δηλαδή την Εξόρυξη Εκπαιδευτικών Δεδομένων, ασχολείται με την ανάπτυξη τεχνικών-μεθόδων σε δεδομένα που προέρχονται αποκλειστικά από την εκπαιδευτική διαδικασία. Η ΕΕΔ επικεντρώνεται στη διερεύνηση νέων τεχνολογιών(για παράδειγμα νέοι αλγόριθμοι) και στην αναζήτηση νέων μοτίβων στα δεδομένα(Romero,Ventura,2020). Από την άλλη πλευρά, η Μαθησιακή Αναλυτική χρησιμοποιεί τις ήδη γνωστές μεθόδους για να επηρεάσει θετικά το Εκπαιδευτικό Σύστημα. Αναλυτικότερα, η ΜΑ αξιοποιεί τα εκπαιδευτικά δεδομένα, με σκοπό να καταλάβει ο καθηγητής και ο ερευνητής τις δραστηριότητες που διενεργούνται κατά τη διάρκεια της διδασκαλίας. Συνεπώς, λαμβάνονται υπόψιν όλες οι εναλλακτικές επιλογές-διορθώσεις στον τομέα της εκπαίδευσης και προτείνεται στους αρμόδιους φορείς το μοντέλο με τη μεγαλύτερη ακρίβεια/πιθανότητα επιτυχίας. Τέλος, η μεγαλύτερη ίσως διαφορά της με την ΕΕΔ είναι η εστίαση της σε ένα ευρύτερο φάσμα, συνεργαζόμενη με τομείς όπως η Ψυχολογία και η Κοινωνιολογία(Bienkowski,Feng,Means,2012).

Συμπερασματικά στόχο του θεωρητικού μέρους της συγκεκριμένης εργασίας αποτελεί η εισαγωγή στις έννοιες της Εξόρυξης Εκπαιδευτικών Δεδομένων και της Μαθησιακής Αναλυτικής. Ειδικότερα, θα προβούμε σε αναφορά αλλά και αναλυτικό σχολιασμό, με τη βοήθεια ερευνητικών-ακαδημαϊκών πηγών που βρίσκονται αναρτημένες στο Διαδίκτυο, των μεθόδων Εξόρυξης Δεδομένων στην Εκπαίδευση. Επίσης, θα περιγράψουμε τις τεχνικές για την επεξεργασία εκπαιδευτικών δεδομένων και θα αναφερθούμε σε όλα τα στάδια της διαδικασίας. Στον τομέα της Μαθησιακής Αναλυτικής, θα καταγράψουμε τόσο τους περιορισμούς και τα ζητήματα που πρέπει να ληφθούν υπόψιν, όσο και τα οφέλη της στην εκπαίδευση.

Εν κατακλείδι, στο Πρακτικό μέρος της εργασίας θα ασχοληθούμε με ένα εργαλείο ΕΕΔ και ΜΑ που υπάρχει δωρεάν στο Διαδίκτυο, και θα παρουσιάσουμε την εμπειρία χρήσης του.

***Θεωρητικό Μέρος***

***Κεφάλαιο 1: Εισαγωγή στην έννοια της εξόρυξης δεδομένων και της μαθησιακής αναλυτικής***

* **Εξόρυξη Δεδομένων**

Όπως αναφέραμε και στην Εισαγωγή της εργασίας, στις μέρες μας σημειώνεται εκρηκτική αύξηση της ποσότητας των δεδομένων που παράγονται. Ως αποτέλεσμα, υπάρχουν μεγάλες ποσότητες δεδομένων αποθηκευμένες κατά κύριο λόγο σε Βάσεις Δεδομένων(ΒΔ). Το μέγεθος της πληροφορίας, σε όποια μορφή και αν βρίσκεται(έγγραφα, διάγραμμα, φωτογραφία, βίντεο), δε μπορεί να κατανοηθεί και να αναλυθεί από τον ανθρώπινο εγκέφαλο. Οπότε, όπως γίνεται εύκολα αντιληπτό, υπήρξε ανάγκη ανακάλυψης γνώσης σε ΒΔ. Η συγκεκριμένη ανάγκη αποτέλεσε την αιτία για την αλματώδη αύξηση του επιστημονικού ενδιαφέροντος στον χώρο της Εξόρυξης Δεδομένων(ΕΔ). Αξίζει να σημειωθεί ότι είναι απαραίτητος ο σχολιασμός και η ανάλυση της ΕΔ, πριν να ασχοληθούμε με την Εξόρυξη Εκπαιδευτικών Δεδομένων.

Ο στόχος της ΕΔ είναι να εντοπίσει χρήσιμες και προηγουμένως άγνωστες πληροφορίες και μοτίβα στα υπάρχοντα δεδομένα(Chung,Gray,1999). Συνήθως αναφέρεται και με τον όρο Ανακάλυψη Γνώσης σε Βάσεις Δεδομένων(KDD). Σημειώνεται ότι η ΕΔ είναι ένα σημαντικό βήμα στη γενικότερη διαδικασία της Ανακάλυψης Γνώσης(Peacock,1998). Αναλυτικότερα, τα βήματα που προηγούνται της ΕΔ είναι τα εξής: Καθαρισμός, Μετασχηματισμός και Επιλογή Δεδομένων. Αναλυτικότερα, κατά τον καθαρισμό των δεδομένων πραγματοποιούνται διάφορες σημαντικές ενέργειες, όπως η διαχείριση ελλιπών τιμών, ο εντοπισμός ακραίων τιμών και η αντιμετώπιση του θορύβου. Οι συγκεκριμένες ενέργειες είναι απολύτως απαραίτητες για να μπορέσουμε να εξάγουμε αποτελέσματα. Όσον αφορά το μετασχηματισμό των δεδομένων, μορφή του οποίου είναι και η Κανονικοποίηση, είναι κατά κάποιο τρόπο προετοιμασία των δεδομένων, ώστε να τα εισάγουμε ως είσοδο σε κάποια μέθοδο Εξόρυξης Δεδομένων. Περιλαμβάνει τη διακριτοποίηση και τη γενικότερη μείωση των διαστάσεων των δεδομένων. Ενώ, στο βήμα της επιλογής των δεδομένων, ανάλογα τη μορφή του προβλήματος που αντιμετωπίζουμε, επιλέγουμε ποια από τα δεδομένα του συνόλου που έχουμε συλλέξει θα χρησιμοποιηθούν στα επόμενα στάδια. Λαμβάνουμε υπόψιν μας τη συσχέτιση των χαρακτηριστικών, και κρατάμε μόνο τα απαραίτητα, δηλαδή αυτά που μας δίνουν σημαντικές πληροφορίες. Σύμφωνα με papers που αναγνώσαμε, σε αυτό το βήμα χρησιμοποιούνται τεχνικές επιλογής χαρακτηριστικών που περιέχουν Μηχανική Μάθηση. Ενώ, έπειτα από την ΕΔ διενεργείται η Αποτίμηση των δεδομένων και η Αναπαράσταση τους(διαφάνειες καθηγητή Μαστοροκώστα στο μάθημα Εξόρυξη Δεδομένων). Η Αποτίμηση αποτελεί την μέτρηση της ποιότητας των πληροφοριών που εξάχθηκαν κατά τη διαδικασία της Εξόρυξης Δεδομένων. Τέλος, η Αναπαράσταση των δεδομένων, παρουσιάζει με τη βοήθεια κυρίως γραφημάτων τα αποτελέσματα της Εξόρυξης Δεδομένων. Μέσω της οπτικοποίησης των αποτελεσμάτων καταλαβαίνει ο απλός χρήστης της εφαρμογής, αλλά και κάποιος επαγγελματίας του χώρου ή ερευνητής σημαντικές πληροφορίες που αγνοούσε ή που δεν πίστευε ότι θα μπορούσαν ποτέ να αναπαρασταθούν με απλό τρόπο, ώστε να τα κατανοήσει πλήρως.

Δίνεται μεγάλη έμφαση στην απόδοση της διαδικασίας της ΕΔ, καθώς τα συστήματα ΕΔ διαχειρίζονται τεράστιο όγκο δεδομένων που προέρχεται από ετερογενείς πηγές δεδομένων. Επίσης, πρέπει να αναφερθεί ότι η ΕΔ είναι χρήσιμη-απαραίτητη για αρκετούς κλάδους όπως η Οικονομία, η Εκπαίδευση, ο Αθλητισμός, η Υγεία και πολλοί άλλοι. Ως επιστημονικό πεδίο «συνορεύει» με τη Στατιστική, τη Μηχανική Μάθηση, την Επιστήμη των Δεδομένων και χρησιμοποιεί τεχνικές που βρίσκονται στο φάσμα της Τεχνητής Νοημοσύνης.

Σε πολλές περιπτώσεις έχει παρατηρηθεί ότι δεν είναι εκ των προτέρων γνωστή η χρήσιμη πληροφορία. Για αυτό το λόγο είναι απαραίτητος ο συνδυασμός διαφορετικών λειτουργιών Εξόρυξης Δεδομένων. Ειδικότερα, η επιστημονική κοινότητα έχει χωρίσει τις λειτουργίες/μοντέλα σε δύο μεγάλες κατηγορίες: Περιγραφικά και Προγνωστικά. Με τον όρο περιγραφικά μοντέλα αναφερόμαστε στην αναγνώριση προτύπων που υπάρχουν στα δεδομένα. Εξετάζει δηλαδή τις ιδιότητες των δεδομένων. Αντίθετα, τα μοντέλα πρόβλεψης(προγνωστικά) επεξεργάζονται τα δεδομένα ακολουθώντας συγκεκριμένη μεθοδολογία με στόχο την παραγωγή προβλέψεων. Παράδειγμα της συγκεκριμένης κατηγορίας αποτελεί μια εφαρμογή εκτίμησης της κίνησης στους δρόμους μιας πόλης, χρησιμοποιώντας δεδομένα που συλλέγονται από κάμερες και αισθητήρες. Από την άλλη πλευρά παράδειγμα συστήματος περιγραφής είναι ένα ηλεκτρονικό ρολόι(smart watch) το οποίο καταγράφει σημαντικές πληροφορίες για την υγεία του ατόμου που το φορά, αλλά και για τη γενικότερη δραστηριότητα του. Υπάρχουν αρκετά μοντέλα περιγραφής αλλά και πρόβλεψης. Σε αυτό το σημείο απλά θα αναφερθούν, ενώ ο αναλυτικός σχολιασμός τους θα πραγματοποιηθεί στο Κεφάλαιο 2:. Αρχικά, η Κατηγοριοποίηση και όλες οι υποκατηγορίες της, αποτελεί την τοποθέτηση μιας εισόδου σε μία από τις προκαθορισμένες κατηγορίες. Επίσης, η Πρόβλεψη αναφέρεται στην εκτίμηση μιας μελλοντικής τιμής, λαμβάνοντας υπόψιν προηγούμενες τιμές και διάφορους παράγοντες. Παράλληλα, η Παλινδρόμηση εστιάζει στην πρόβλεψη σε συστήματα συνεχούς τιμής. Επιπροσθέτως, αρκετά δημοφιλής είναι και η Συσταδοποίηση, η οποία αφορά την ομαδοποίηση των δεδομένων σε συστάδες με βάση την ομοιότητα τους. Τα πρώτα 3 μοντέλα που αναφέραμε αποτελούν παραδείγματα επιβλεπόμενης μάθησης, ενώ η Συσταδοποίηση ανήκει στην μη-επιβλεπόμενη μάθηση. Αξίζει να σημειώσουμε ότι η Κατηγοριοποίηση είναι μία αρκετά μεγάλη κατηγορία με πολλές υποκατηγορίες, που δεν ανήκουν όλες σε τεχνικές επιβλεπόμενης μάθησης(για παράδειγμα οι Κανόνες Συσχέτισης είναι μοντέλο μη-επιβλεπόμενης μάθησης). Με τον όρο επιβλεπόμενη μάθηση αναφερόμαστε σε έναν τύπο μηχανικής μάθησης όπου το μοντέλο εκπαιδεύεται χρησιμοποιώντας ένα dataset που εκτός από τις εισόδους περιλαμβάνει και τις επιθυμητές εξόδους, και τις συγκρίνει με τις πραγματικές εξόδους του συστήματος. Κατά κάποιο τρόπο γίνεται επίβλεψη του τρόπο μάθησης του συστήματος, και διορθώνεται μέσω της γνώσης που έχει αποκτήσει μέσα από τη διαδικασία εκπαίδευσης του.

Στη συνέχεια, θα αναφέρουμε 3 βασικά στάδια της Εξόρυξης Δεδομένων: Περιγραφή Μοντέλου, Αξιολόγηση Μοντέλου και Αλγόριθμος Αναζήτησης(Fayyad,1996). Αναλυτικότερα, στο πρώτο στάδιο ορίζεται το πρόβλημα και οι στόχοι του μοντέλου. Επίσης, λαμβάνει χώρα και η επιλογή των κατάλληλων δεδομένων, αλλά και η επεξεργασία τους. Έπειτα, γίνεται η Αξιολόγηση του μοντέλου, δηλαδή επιλέγεται το κατάλληλο μοντέλο για το συγκεκριμένο πρόβλημα και πραγματοποιείται η εκπαίδευση του με τα επιλεγμένα δεδομένα. Τέλος, εφαρμόζονται αλγόριθμοι ΕΔ(Κατηγοριοποίησης, Συσταδοποίησης, …). Ακολουθεί η αξιολόγηση των αποτελεσμάτων με τη βοήθεια κάποιας μετρικής(ακρίβεια, ανάκληση, F1-score). Το τελευταίο βήμα του τρίτου σταδίου είναι η βελτιστοποίηση του μοντέλου.

Αναντίρρητα, επίκαιρο και πολύ σημαντικό είναι το ζήτημα της ασφάλειας των δεδομένων. Υπάρχουν νόμοι με κεντρικό τους θέμα τα δεδομένα, που προστατεύουν πολύ σημαντικά ανθρώπινα δικαιώματα, όπως η Ιδιωτικότητα. Με μια πρώτη ανάγνωση, σκεφτόμαστε πως το παραπάνω είναι ερευνητικό ζήτημα της Ασφάλειας Πληροφοριακών Συστημάτων. Ωστόσο, αντανακλά στην Εξόρυξη Δεδομένων, καθώς τα δεδομένα είναι αναπόσπαστο κομμάτι της. Επίσης, η ΕΔ ασχολείται και με τον εντοπισμό διαφόρων απατών και σπάνιων δεδομένων. Τα παραπάνω ανήκουν στην Ανάλυση των ακραίων δεδομένων η οποία αποτελεί πολύ σημαντικό τμήμα της Εξόρυξης Δεδομένων. Εν συντομία, ως ακραία χαρακτηρίζονται τα δεδομένα που δε συμφωνούν με τη γενική εικόνα/συμπεριφορά των υπόλοιπων δεδομένων(διαφάνειες καθηγητή Μαστοροκώστα στο μάθημα Εξόρυξη Δεδομένων).

* **Εξόρυξη Εκπαιδευτικών Δεδομένων**

Προηγήθηκε αναλυτικός σχολιασμός των κυριότερων σημείων της Εξόρυξης Δεδομένων. Οπότε, βρισκόμαστε στο κατάλληλο σημείο για να εντάξουμε στο συγκεκριμένο έγγραφο έννοιες που προέρχονται από το ερευνητικό πεδίο της Εξόρυξης Εκπαιδευτικών Δεδομένων. Καταρχάς, η ΕΕΔ αποτελεί ένα ερευνητικό πεδίο που εστιάζει στην εξαγωγή ενδιαφέρουσας γνώσης στον τομέα της εκπαίδευσης(Trung,Son,Tung,Anh,Lam,2023). Μετά από σχετική έρευνα καταλήξαμε στο γεγονός ότι συνεργάζεται με τα εξής πεδία: Εκπαίδευση, Ηλεκτρονική Μάθηση, Αλληλεπίδραση ανθρώπου-υπολογιστή, Ψυχολογία και αρκετά ακόμη. Αξίζει να αναφερθεί ότι ο τομέας που αναλύουμε εμφανίζει σημαντική άνθηση τα τελευταία χρόνια. Χαρακτηριστικά, κάθε χρόνο διοργανώνονται συνέδρια με κεντρικό θέμα την ΕΕΔ, ενώ πραγματοποιείται συγγραφή πολυάριθμων papers από διακεκριμένους επιστήμονες. Σύμφωνα με τους ειδικούς, η παραπάνω άνθιση έλαβε χώρα εξαιτίας της σημασίας που έχει διαχρονικά η Εκπαίδευση για το άτομο, και γενικότερα για την κοινωνία.

Για να γίνουμε πιο συγκεκριμένοι, στόχο της Εξόρυξης Εκπαιδευτικών Δεδομένων αποτελεί η βελτίωση της εκπαιδευτικής διδασκαλίας. Ειδικότερα, γίνονται έρευνες με σκοπό να κατανοήσει η επιστημονική κοινότητα τον τρόπο με τον οποίο μαθαίνει ο μαθητής/φοιτητής και πως μπορεί να βελτιωθεί η ψυχολογία του μέσα στην τάξη(ισχύει και για την εξ αποστάσεως εκπαίδευση). Επίσης, επιδιώκουν να προβλέψουν τη μελλοντική μαθησιακή συμπεριφορά των μαθητών και των φοιτητών με τη βοήθεια σχετικών μοντέλων(Bienkowski,Feng,Means,2012). Επομένως, η ΕΕΔ χωρίζει(κατηγοριοποιεί) τις ενέργειες που γίνονται από έναν μαθητή σε σχετικές και μη με τη μαθησιακή πρόοδο του. Κατά κάποιο τρόπο, σχεδιάζεται μια στρατηγική που στοχεύει στην ατομική βελτίωση του κάθε μαθητή ξεχωριστά, όσο και στη βελτίωση της γενικότερης εκπαιδευτικής διαδικασίας. Χωρίς αμφιβολία, δε νοείται Εξόρυξη (Εκπαιδευτικών) Δεδομένων χωρίς αναφορά σε δεδομένα. Στη σημερινή εποχή, οι ερευνητές ασχολούνται ενεργά με τη δημιουργία εφαρμογών/συστημάτων εκπαιδευτικού ενδιαφέροντος, όπου πραγματοποιείται απευθείας συλλογή δεδομένων, μέσω της αλληλεπίδρασης των μαθητών/φοιτητών με τη γραφική διεπαφή(user interface). Επίσης, γίνονται προσπάθειες για την ανάλυση της συμπεριφοράς του μαθητή κατά τη διάρκεια εξ αποστάσεως μαθήματων, και η απαραίτητη σύγκριση με τη δια ζώσης διδασκαλία. Τέλος, είναι αναγκαία η αναφορά στην αυστηρότατη πολιτική που εφαρμόζει η Ευρωπαϊκή Ένωση, αλλά και οι υπόλοιποι οργανισμοί, στη διαδικασία της συλλογής εκπαιδευτικών δεδομένων που αναφέραμε παραπάνω. Τα συγκεκριμένα δεδομένα χαρακτηρίζονται ως ευαίσθητα από την επιστημονική κοινότητα, και υπάρχουν πολυάριθμοι κανόνες που σχετίζονται με τη συλλογή, αλλά και με τη χρήση τους. Το παραπάνω ζήτημα κυριάρχησε στα συνέδρια της ΕΕΔ, έπειτα από προτάσεις για είσοδο της Βαθιάς Μάθησης(Deep Learning) στην εκπαίδευση. Εν συντομία, η Βαθιά Μάθηση βρίσκεται στο φάσμα της Μηχανικής Μάθησης και χρησιμοποιεί πολυστρωματικά νευρωνικά δίκτυα για την επίλυση σύνθετων προβλημάτων(Trung,Son,Tung,Anh,Lam,2023). Η Βαθιά Μάθηση αποτελεί ένα από τα πιο ενδιαφέροντα πεδία τη συγκεκριμένη χρονική στιγμή και είναι ιδιαίτερα αποτελεσματική. Χρησιμοποιήσαμε το παράδειγμα της Βαθιάς Μάθησης για να γίνει αντιληπτό το γεγονός ότι οι επιστήμονες δεν επιλέγουν τεχνικές ανάλυσης εκπαιδευτικών δεδομένων που εμφανίζουν τη μέγιστη αποτελεσματικότητα, αλλά είναι υποχρεωμένοι να υιοθετήσουν τους νόμους που έχουν θεσπιστεί.

Η ΕΕΔ εφαρμόζει τεχνικές Εξόρυξης Δεδομένων έχοντας 3 κύριους αποδέκτες. Αρχικά, κεντρικό πρόσωπο όλων των δράσεων αποτελούν οι μαθητές. Έπειτα, οι καθηγητές και στο τέλος το Υπουργείο Παιδείας. Οι επιστήμονες αναφέρουν ως απαραίτητη τη συνεργασία της ερευνητικής κοινότητας με τους φορείς της εκπαίδευσης, έχοντας ως σκοπό τη διαφοροποίηση του περιεχομένου του μαθήματος και την ταξινόμηση των μαθητών σε ομάδες με βάση τις ανάγκες του ή κοινά χαρακτηριστικά που διαθέτουν/εμφανίζουν(Romero,Ventura,2007). Ακόμη, οι επιστήμονες τονίζουν τη σημασία της ΕΕΔ στα πανεπιστήμια. Σημειώνεται ότι θα πραγματοποιηθεί αναλυτικός σχολιασμός των μεθόδων-τεχνικών ΕΔ που χρησιμοποιούνται στην εκπαίδευση, σε επόμενο Κεφάλαιο της εργασίας.

* **Μαθησιακή Αναλυτική**

Όπως αναφέραμε και στην Εισαγωγή της παρούσας εργασίας, η Μαθησιακή Αναλυτική(ΜΑ) αποτελεί ξεχωριστό ερευνητικό πεδίο, και σημειώνει σημαντική αύξηση σε αριθμό δημοσιεύσεων την τελευταία δεκαετία. Ως ορισμός ανακτήθηκε ο εξής(στα περισσότερα papers είναι παρόμοιος): Κατανόηση και βελτίωση της διαδικασίας της μάθησης. Επίσης, έχει χαρακτηριστεί ως κατανόηση της μαθησιακής συμπεριφοράς, μέσω της αλληλεπίδρασης μαθητή/φοιτητή με τον καθηγητή και με τα δεδομένα που δημιουργούνται(Wong,Chang,2018). Αναντίρρητα, η εξ αποστάσεως εκπαίδευση παρέχει μεγάλο όγκο δεδομένων στους επιστήμονες που ασχολούνται ενεργά με τη ΜΑ. Η ΜΑ μέσω μοντέλων και τεχνικών ενημερώνει εκπαιδευτικούς και φορείς, όπως το Υπουργείο Παιδείας κάθε κράτους, για την πορεία που διαγράφει ο μαθητής/φοιτητής σε συγκεκριμένο χρονικό διάστημα Όσον αφορά ο πανεπιστήμιο, που χρησιμοποιούνται σε μεγαλύτερο βαθμό από τα σχολεία ηλεκτρονικές εφαρμογές μάθησης, η συλλογή δεδομένων είναι ευκολότερη διαδικασία. Επίσης, τα μαζικά ανοιχτά μαθήματα που πραγματοποιούνται μέσω διαδικτύου αποτελούν αρκετά σημαντική πηγή δεδομένων για την επιστημονική κοινότητα (Lee,Cheung,Kwok,2020).

Σε αυτό το σημείο θα αναφερθούν συγκεκριμένα παραδείγματα δράσης της Μαθησιακής Αναλυτικής. Ειδικότερα, όσον αφορά τα πανεπιστήμια, γίνεται πρόβλεψη του αριθμού των φοιτητών που θα παρακολουθήσουν ένα συγκεκριμένο μάθημα ενός τμήματος σε όλη τη διάρκεια του εξαμήνου(Dietz-Uhler,Hurn,2013). Επίσης, πραγματοποιείται πρόβλεψη για την επιτυχία ή όχι συγκεκριμένου φοιτητή(δημιουργία προφίλ και παροχή πολυδιάστατων δεδομένων) στις τελικές εξετάσεις ενός μαθήματος που έχει δηλώσει στην αρχή του εξαμήνου. Επιπροσθέτως, δύναται να χρησιμοποιηθεί από την πλευρά του Υπουργείου Παιδείας και της Κυβέρνησης κάποιας χώρας να γίνει δίκαιη χρηματοδότηση των πανεπιστημίων. Αξίζει να σημειωθεί ότι η Μαθησιακή Αναλυτική παρέχει εξατομικευμένη μάθηση, δηλαδή ξεχωριστό τρόπο μετάδοσης της γνώσης για κάθε μαθητή/φοιτητή ξεχωριστά(Greller,Drechsler,2012). Ακόμη, γίνεται αναφορά σε δημοσιεύσεις με θέμα τη ΜΑ, ότι είναι δυνατή η πρόβλεψη του χρονικού διαστήματος που πρέπει να μεσολαβήσει για να προχωρήσει ο καθηγητής στη διδασκαλία της επόμενης μαθησιακής ενότητας, και σε τι βαθμό οι μαθητές έκαναν «κτήμα» τους την προηγούμενη ενότητα. Τέλος, είναι δυνατή η παροχή συμβουλών στους φοιτητές σχετικά με την επιλογή συγκεκριμένων μαθημάτων, των οποίων το επιστημονικό πεδίο που βρίσκονται θα σημειώσει άνθιση στο μέλλον, αλλά και μαθημάτων που ταιριάζουν στις δεξιότητες του συγκεκριμένου φοιτητή.

Αφού έγινε η απαραίτητη αναφορά στον ορισμό της έννοιας Μαθησιακή Αναλυτική, και ακολούθησαν συγκεκριμένα παραδείγματα χρήσης της ΜΑ σε σχολεία και πανεπιστήμια, θα ακολουθήσει η αναφορά σε συγκεκριμένες μορφές δεδομένων που απαιτεί η διαδικασία για να χαρακτηριστεί επιτυχημένη. Καταρχάς, είναι γνωστό ότι τα πανεπιστήμια έχουν στη διάθεσή τους μεγάλο όγκο δεδομένων. Αναλυτικότερα, στο παρελθόν έχει χρησιμοποιηθεί σύστημα που καταγράφει τη συχνότητα σύνδεσης του φοιτητή σε μια πλατφόρμα παρόμοια με το eclass, τη χρονική διάρκεια της σύνδεσης του σε αυτή, τον αριθμό των διαλέξεων που παρευρέθηκε σε ένα συγκεκριμένο εξάμηνο για ένα μάθημα και τους βαθμούς των εργασιών που υπέβαλλε ο φοιτητής, με στόχο να προβλεφθεί ο βαθμός του στην τελική εξέταση(Lange,Huston,2002). Όπως γίνεται αντιληπτό, για να βγάλει στην έξοδο του χρήσιμα αποτελέσματα ένα σύστημα ΕΕΔ και ΜΑ, πρέπει να υπάρξει συνεργασία μεταξύ σχολείου/πανεπιστημίου, καθηγητών, Υπουργείου Παιδείας και φυσικά συμμετοχή των μαθητών/φοιτητών. Εάν κάποιος από τους παραπάνω δε συμμετάσχει στη διαδικασία της παραγωγής των δεδομένων, τότε η επεξεργασία και η ανάλυση τους γίνεται πιθανότατα χωρίς νόημα. Επιπροσθέτως, το 2010 διεξήχθη μια μελέτη για τη ΜΑ, όπου ο καθηγητής συνέστησε στους μαθητές του να συμμετέχουν σε ένα forum συζήτησης για το συγκεκριμένο μάθημα. Μέσα από τη συζήτηση μεταξύ των μαθητών, αλλά και με τον καθηγητή, οι περισσότεροι μαθητές προέβλεψαν επιτυχώς τον τελικό βαθμό τους στο μάθημα(Felakmasir,Jafar,2010). Αξίζει να αναφερθεί το γεγονός ότι τα ερωτηματολόγια που διανέμονται σε φοιτητές από τους καθηγητές τους και γενικότερα από φορείς του πανεπιστημίου, βοηθούν αρκετά στην κατανόηση των συνθηκών κάτω από τις οποίες διενεργείται η διαδικασία του μαθήματος. Επίσης, μέσω των ερωτηματολογίων γίνεται και κατά κάποιο τρόπο αξιολόγηση των καθηγητών, των βιβλίων που παρέχονται στην αρχή του εξαμήνου και γενικά στις συνθήκες φοίτησης. Ακόμη, τα ερωτηματολόγια μπορούν να έχουν και μια λειτουργία πέρα από τα όρια του αμιγώς μαθησιακού περιβάλλοντος. Δηλαδή, συγκεκριμένες ερωτήσεις στοχεύουν στην απόκτηση πληροφοριών που αφορούν τη ψυχολογία του φοιτητή και το βαθμό κοινωνικοποίησης του. Με αυτό τον τρόπο, παρέχεται βοήθεια σε φοιτητές που τη χρειάζονται σε σοβαρά θέματα.

Τέλος, θα σχολιάσουμε εν συντομία τα στάδια-βήματα που έχουν οριστεί από την επιστημονική κοινότητα ως διαδικασία της Μαθησιακής Αναλυτικής. Αρχικά, χωρίζεται σε 5 στάδια: καταγραφή, αναφορά, πρόβλεψη, πράξη και βελτιστοποίηση(Cambell,Oblinger,2012). Οι συγκεκριμένοι επιστήμονες τονίζουν ότι υπάρχει μεγάλη εξάρτηση μεταξύ των 5 βημάτων. Ως καταγραφή ορίζουν τη συλλογή και την αποθήκευση δεδομένων που αφορούν τη ΜΑ και λαμβάνουν χώρα στη μάθηση. Στη συνέχεια, πραγματοποιείται η επεξεργασία των δεδομένων. Στο τρίτο βήμα, γίνονται προβλέψεις με βάση τα δεδομένα που αποκτώνται για το μαθητή/φοιτητή, για τον εκπαιδευτικό, για συγκεκριμένα μαθήματα και γενικότερα για σημαντικά εκπαιδευτικά ζητήματα. Αναφέραμε αρκετά παραδείγματα πρόβλεψης σε προηγούμενο σημείο της εργασίας. Στο επόμενο βήμα, και αφού αναλυθούν οι προβλέψεις του προηγούμενου βήματος, σχεδιάζεται η στρατηγική που θα ακολουθηθεί για την αντιμετώπιση των προβλημάτων που φανερώθηκαν. Αναμφίβολα, το συγκεκριμένο στάδιο βασίζεται στη συνεργασία των καθηγητών με συγκεκριμένους φορείς του εκάστοτε πανεπιστημίου και του Υπουργείου Παιδείας. Οι Cambell και Oblinger όρισαν ως τελευταίο βήμα της διαδικασίας τη βελτιστοποίηση. Στόχος της είναι η βελτίωση των αποτελεσμάτων που εξάγουν τα προηγούμενα βήματα. Ως παράδειγμα δίνεται η αύξηση της ακρίβειας των αποτελεσμάτων, δηλαδή η διόρθωση του βήματος 3(πρόβλεψη).

* **ΕΔ και ΜΑ**

Τα 2 βασικά ερευνητικά πεδία που ασχολούμαστε στην εργασία είναι η ΕΔ και η ΜΑ. Η ανάπτυξη και των δύο ξεκίνησε σχετικά πρόσφατα, τις τελευταίες δεκαετίες ειδικότερα, και σχετίζεται με τη γενικότερη Ψηφιακή Επανάσταση. Καταρχάς, μέσω του παραπάνω αναλυτικού σχολιασμού τους, κατανοήσαμε ότι έχουν κοινή βάση. Χρησιμοποιούν δηλαδή δεδομένα που προέρχονται από την εκπαίδευση. Αναμφίβολα, έχουν και κοινό στόχο, τη βελτίωση του Εκπαιδευτικού Συστήματος. Για την επίτευξη του συγκεκριμένου στόχου, απαιτείται η συνεργασία τους. Θα μπορούσαμε να πούμε ότι αλληλοσυμπληρώνονται. Χαρακτηριστικά, συνδυάζονται για να προσφέρουν αξιόπιστες αναλύσεις στον τομέα της εκπαίδευσης. Αναλυτικότερα, για τη βελτίωση της διδασκαλίας, η ΕΕΔ καλείται να αναλύσει τα δεδομένα που σχετίζονται με την πορεία των μαθητών, τις γνώσεις που μεταλαμπαδεύει ο καθηγητής, τις εξετάσεις των μαθητών και με αρκετούς ακόμη παράγοντες. Η ΜΑ με τη σειρά της εφαρμόζει μοντέλα για την πρόβλεψη των επιδόσεων μαθητών-καθηγητών και συστήματος εκπαίδευσης, ενώ χρησιμοποιεί και εξατομικευμένα μοντέλα πρόβλεψης και αναγνώρισης προτύπων. Η ΜΑ δηλαδή χρησιμοποιεί τεχνικές ΕΔ σε δεδομένα που προέρχονται από τον τομέα της εκπαίδευσης στη διαδικασία της λήψης αποφάσεων.

Επιπροσθέτως, στο κέντρο και των 2 εννοιών βρίσκεται ο μαθητής/φοιτητής, ο καθηγητής και το Εκπαιδευτικό Σύστημα. Χωρίς αμφιβολία, όπως αναφέραμε και στην Εισαγωγή της εργασίας, εμφανίζουν και ορισμένες διαφορές, οπότε δεν πρέπει να συγχέονται.. Ειδικότερα, η ΕΕΔ είναι αμιγώς εστιασμένη στην Πληροφορική και την Τεχνητή Νοημοσύνη. Χρησιμοποιεί μεθόδους και τεχνικές ΕΔ σε δεδομένα που προέρχονται από την εκπαίδευση. Αντίθετα, η ΜΑ επικεντρώνεται στις προβλέψεις και τις αναλύει με άξονα και τις Ψυχολογία, Κοινωνιολογία εκτός από την Πληροφορική.

***Κεφάλαιο 2:Μέθοδοι εξόρυξης δεδομένων που εφαρμόζονται στον τομέα της εκπαίδευσης***

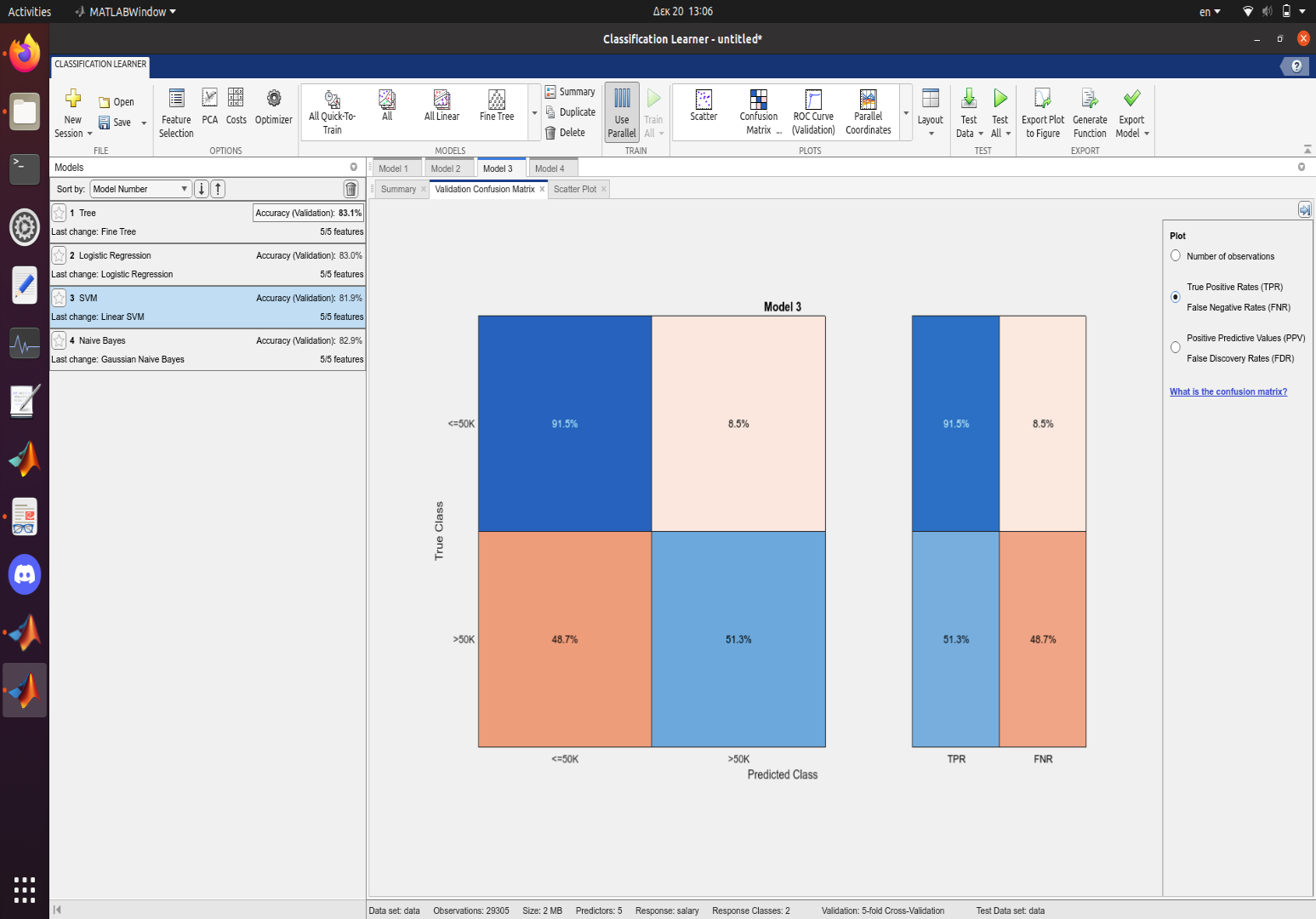
Τόσο στην Εισαγωγή της εργασίας, όσο και στο πρώτο Κεφάλαιο έχουμε ορίσει την έννοια Εξόρυξη Δεδομένων, την έχουμε περιγράψει αναλυτικά και έχουμε αναφέρει κάποιες από τις μεθόδους της. Σε αυτό το σημείο ακολουθεί διεξοδική ανάλυση συγκεκριμένων δημοφιλών μεθόδων εξόρυξης δεδομένων που ανήκουν στις κατηγορίες των περιγραφικών μοντέλων, αλλά και των μοντέλων πρόβλεψης. Ενώ, στη συνέχεια θα εξηγήσουμε, έπειτα από ανάγνωση αρκετών ερευνητικών εργασιών, πως οι παραπάνω μέθοδοι εφαρμόζονται στην πράξη στον τομέα της εκπαίδευσης.

* **Κατηγοριοποίηση**

Καταρχάς, η Κατηγοριοποίηση είναι μια δημοφιλής τεχνική που χρησιμοποιείται αρκετά σε εφαρμογές Εξόρυξης Δεδομένων και Μηχανικής Μάθησης. Ασχολείται με την ταξινόμηση αντικειμένων σε διάφορες κατηγορίες με βάση κάποια από τα χαρακτηριστικά τους. Σε κάθε εφαρμογή που χρησιμοποιείται η συγκεκριμένη τεχνική, υπάρχει ένα σύνολο δεδομένων ή σωστότερα ένα σύνολο από αντικείμενα/στιγμιότυπο. Κάθε στιγμιότυπο χαρακτηρίζεται από 2 μεταβλητές: x και y. Το x αποτελεί το σύνολο τιμών των χαρακτηριστικών, ενώ το y είναι ένα κατηγορικό χαρακτηριστικό. Αξίζει να τονιστεί ότι ένα μοντέλο κατηγοριοποίησης είναι η αναπαράσταση της σχέσης μεταξύ των x-y(Εισαγωγή στην Εξόρυξη Δεδομένων,Tan,Steinbach,Karpatne,Kumar). Οπότε, κάθε στιγμιότυπο περιέχει ένα σύνολο από χαρακτηριστικά. Ένα από τα χαρακτηριστικά είναι και η κλάση. Το μοντέλο μπορεί να αναπαρασταθεί με πολλές τεχνικές. Κάποιες από αυτές αποτελούν και πολύ δημοφιλή ερευνητικά πεδία στις μέρες μας και εξελίσσονται συνεχώς. Αυτό που ισχύει για όλες τις τεχνικές είναι ότι για να ταξινομηθεί ένα στιγμιότυπο στην σωστή κατηγορία πρέπει να ισχύει ότι f(x)=y , όπου f μία συνάρτηση στόχου.

Παραπάνω αναφέραμε ότι οι μέθοδοι Εξόρυξης Δεδομένων χωρίζονται σε 2 μεγάλες κατηγορίες: περιγραφικά και μοντέλα πρόβλεψης. Τη σημαντικότητα της Κατηγοριοποίησης τη φανερώνει το γεγονός ότι ανήκει και στις 2 κατηγορίες. Αναλυτικότερα, χρησιμοποιείται ως μοντέλο πρόβλεψης για να ταξινομεί στιγμιότυπα χωρίς ετικέτα. Από την άλλη πλευρά, χρησιμοποιείται και ως μοντέλο της 2ης κατηγορίας, όταν προσδιορίζει τα χαρακτηριστικά(features) τα οποία διακρίνουν τα αντικείμενα που προέρχονται από διαφορετικές κατηγορίες. Στα papers που αναγνώσαμε σημειώνεται ότι το μυστικό στην Κατηγοριοποίηση είναι η εύρεση του ιδανικού συνδυασμού χαρακτηριστικών. Πρέπει να αναφέρουμε ότι τα περισσότερα σύνολα δεδομένων περιέχουν αρκετά μεγάλο αριθμό χαρακτηριστικών, που δεν υπάρχει ούτε η δυνατότητα οπτικοποίησης τους, για να αντιληφθούμε τη μορφή τους ή τι προσφέρουν στην εκάστοτε εργασία που ασχολούμαστε. Έτσι, καταλαβαίνουμε το πόσο καθοριστικό είναι το στάδιο της Προεπεξεργασίας που αναλύσαμε σε προηγούμενο Κεφάλαιο. Δηλαδή, όταν έρθουμε σε επαφή με ένα πολυδιάστατο σύνολο δεδομένων πρέπει να αφαιρεθούν χαρακτηριστικά που δεν είναι χρήσιμα για το πρόβλημα μας ή που ο συνδυασμός τους με άλλα χαρακτηριστικά προκαλεί περαιτέρω προβληματισμό. Επίσης, πρέπει να πραγματοποιηθεί ανάλυση της σχετικότητας τους και μετασχηματισμοί(όπως η κανονικοποίηση) όπου χρειάζεται(διαφάνειες καθηγητή Μαστοροκώστα στο μάθημα Εξόρυξη Δεδομένων).

Άρα, στόχο της μεθόδου αποτελεί η ταξινόμηση των αντικειμένων σε κατηγορίες με βάση τα χαρακτηριστικά τους. Οπότε, πρώτο βήμα της μεθόδου αποτελεί η εκπαίδευση του μοντέλου. Χρησιμοποιούμε κάποιο εκπαιδευτικό σύνολο με ετικέτες κατηγοριών, ώστε το μοντέλο με το πέρασμα των εποχών να αναγνωρίζει παρόμοια μοτίβα σε κάθε κατηγορία. Έπειτα από την εκπαίδευση του μοντέλου ακολουθεί η αξιολόγηση του. Δηλαδή, η δοκιμή του μοντέλου με νέα δεδομένα, τα οποία δεν έχει ξανασυναντήσει(test set). Αναμφίβολα, το σύνολο εκπαίδευσης και το test set πρέπει να είναι διαφορετικά, καθώς πρέπει το μοντέλο που εξάγεται να μπορεί να προβλέψει τις ετικέτες των στιγμιότυπων που δεν έχει επεξεργαστεί σε προηγούμενο στάδιο. Αν πραγματοποιηθεί αυτό που περιγράφουμε, τότε το μοντέλο Κατηγοριοποίησης έχει γενικευτική ικανότητα και δεν έχει υπερπροσαρμοστεί στα δεδομένα εκπαίδευσης. Υπάρχουν αρκετές μετρικές της απόδοσης ενός μοντέλου Κατηγοριοποίησης. Στο σημείο αυτό θα αναφέρουμε την Ακρίβεια, δηλαδή το κλάσμα του αριθμού των σωστών προβλέψεων του μοντέλου προς το συνολικό αριθμό προβλέψεων. Συχνά εφαρμόζεται και η Ανάκληση, αλλά και το F1-score. Και τα 3 μέτρα που αναφέραμε οπτικοποιούνται μέσω ενός Confusion Matrix.



Η παραπάνω φωτογραφία προέρχεται από εργαστηριακή άσκηση στο μάθημα της Μηχανικής Μάθησης. Έχουμε χρησιμοποιήσει το εργαλείο Classification Learner και εργαζόμαστε στο περιβάλλον Matlab. Στην αριστερή πλευρά της οθόνης μας παρατηρούμε ότι δοκιμάζουμε 4 μοντέλα Κατηγοριοποίησης. Κοιτάζουμε την Ακρίβεια του κάθε μοντέλου για το συγκεκριμένο dataset και επιλέγουμε εκείνο με τη μεγαλύτερη απόδοση.

Άρα καταλαβαίνουμε ότι είναι απαραίτητη η επιλογή του κατάλληλου μοντέλου/αλγορίθμου Κατηγοριοποίησης, ανάλογα το σύνολο εκπαίδευσης που χρησιμοποιούμε και ανάλογα το πρόβλημα στο οποίο έχουμε εστιάσει.

Επιπλέον, για να επιλέξουμε ποιος κατηγοριοποιητής ταιριάζει με το πρόβλημα μας, πρέπει να ελέγξουμε το χρόνο που χρειάζεται η εφαρμογή του μοντέλου, αν μπορεί να χειριστεί τιμές χαρακτηριστικών που λείπουν, θόρυβο και ακραίες τιμές.

Ακολουθούν κάποιες από τις πιο γνωστές τεχνικές Κατηγοριοποίησης:

* **Δένδρα Απόφασης**

Αποτελούν κλασικά δένδρα όπου οι εσωτερικοί κόμβοι τους αντιστοιχούν σε χαρακτηριστικά, ενώ τα φύλλα αντιστοιχούν σε κλάσεις. Επιπροσθέτως, οι ακμές που συνδέουν δύο κόμβους του δένδρου περιέχουν ετικέτες και αποτελούν συνθήκες. Μετά την κατασκευή του δένδρου, το διασχίζουμε από τη ρίζα προς τα φύλλα, με στόχο να κατηγοριοποιήσουμε νέες εγγραφές. Χρησιμοποιούμε σύνολα εκπαίδευσης και ελέγχου που έχουν μορφή πίνακα. Για να δημιουργήσουμε το δένδρο απόφασης πρέπει να προβούμε σε κάποιες ερωτήσεις πάνω στα χαρακτηριστικά του στιγμιότυπου. Μετά από την απάντηση που θα «λάβουμε», ακολουθεί νέα ερώτηση(πηγαίνουμε στον αμέσως επόμενο κόμβο του μονοπατιού που διαμορφώνεται από την αμέσως προηγούμενη απάντηση μας) και αυτή η διαδικασία συνεχίζεται μέχρι να ανακαλύψουμε την κλάση στην οποία ανήκει το συγκεκριμένο στιγμιότυπο.

Αξίζει να σημειωθεί το γεγονός ότι από ένα συγκεκριμένο σύνολο δεδομένων γίνεται να κατασκευαστούν πολλά δένδρα απόφασης. Για να κατασκευάσουμε το δένδρο, αρχικά τοποθετούμε όλα τα στιγμιότυπα σε έναν κόμβο. Στη συνέχεια, χωρίζουμε τα στιγμιότυπα με βάση κάποια συνθήκη. Για παράδειγμα, συνθήκη: Διάβασε ο μαθητής για το διαγώνισμα την προηγούμενη ημέρα; Κατανοούμε ότι οι μαθητές θα χωριστούν σε 2 κατηγορίες, σε αυτούς που διάβασαν την προηγούμενη μέρα και σε αυτούς που δεν το έκαναν. Άρα, σε κάθε κόμβο γίνεται αναδρομική κλήση του παραπάνω βήματος. Από τη στιγμή που θα κατασκευαστεί το δένδρο απόφασης ενδέχεται να ακολουθήσουν ορισμένες βελτιώσεις. Υπάρχουν ωστόσο και κάποια ζητήματα σχεδίασης, όπως ποια θα αποτελέσει τη συνθήκη διαχωρισμού για τη διάσπαση των εγγραφών κάθε κόμβου. Δηλαδή, σε κάθε αναδρομικό βήμα του αλγορίθμου επιλέγεται ένα χαρακτηριστικό για να διαχωρίζει τα δεδομένα εκπαίδευσης που σχετίζονται με τον εκάστοτε κόμβο. Ο διαχωρισμός μπορεί να είναι δυαδικός ή πολλαπλός(οι διασπάσεις είναι τόσες όσες οι διαφορετικές τιμές), ανάλογα με τον τύπο των γνωρισμάτων(δυαδικά ,συνεχή, κατηγορικά). Επίσης, πρέπει να επιλέξουμε πότε θα σταματήσει η επέκταση ενός κόμβου. Ο αλγόριθμος τερματίζει όταν όλα τα στιγμιότυπα του συνόλου ελέγχου κατηγοριοποιηθούν σε κάποια από τις υπάρχουσες κλάσεις. Τέλος, αξίζει να αναφέρουμε τα μέτρα ανομοιογένειας των κόμβων που μας βοηθούν να καταλάβουμε ποιος είναι ο βέλτιστος διαχωρισμός. Αναλυτικότερα, αναφερόμαστε στην εντροπία, σ=το δείκτη Gini και το σφάλμα ταξινόμησης.

* **Κανόνες Συσχέτισης**

Οι Κανόνες Συσχέτισης αποτελούν μια μορφή ανάλυσης δεδομένων που αναζητά συσχετίσεις μεταξύ διαφορετικών χαρακτηριστικών σε ένα σύνολο δεδομένων. Ορισμένες από τις συσχετίσεις που θα προκύψουν ενδέχεται να είναι καθοριστικής σημασίας για το πρόβλημα που αντιμετωπίζουμε, να μας παρέχει ενδιαφέρουσες πληροφορίες που δεν είχαμε σκεφτεί και γενικότερα να βελτιώσει την ποιότητα λήψης αποφάσεων. Είναι της μορφής if[Συνθήκη], then[Απόφαση]. Οι δηλώσεις if-then σχηματίζουν σύνολα, μέσω των οποίων υπολογίζουμε τους κανόνες συσχέτισης. Η συγκεκριμένη μέθοδος συνεργάζεται με τη Μηχανική Μάθηση και με τεχνικές Τεχνητής Νοημοσύνης, με σκοπό να μπορέσουν να αναλυθούν Big Data στις μέρες μας. Συμπερασματικά, ο συγκεκριμένος κατηγοριοποιητής οργανώνει ένα στιγμιότυπο με βάση έναν κανόνα, ο οποίος ενεργοποιείται από το συγκεκριμένο στιγμιότυπο.

Για παράδειγμα αν το 70% των φοιτητών που δηλώνουν το μάθημα Δομές Δεδομένων στο τμήμα Μηχανικών Πληροφορικής και Υπολογιστών του ΠΑΔΑ παρακολουθούν επίσης ένα συγκεκριμένο tutorial στο YouTube, τότε υπάρχει ένα διακριτό μοτίβο στα δεδομένα, ότι οι φοιτητές που δηλώνουν το μάθημα Δομές Δεδομένων του 3ου εξαμήνου συχνά παρακολουθούν ένα συγκεκριμένο tutorial. Καταλαβαίνουμε δηλαδή ότι υπάρχει συσχέτιση. Έπειτα, μπορεί να προκύψει περαιτέρω ανάλυση και σημαντικές πληροφορίες, εάν εξετάσουμε τους βαθμούς της εξεταστικής του συγκεκριμένου ποσοστού για το συγκεκριμένο μάθημα, ή και παρόμοια λάθη τους.

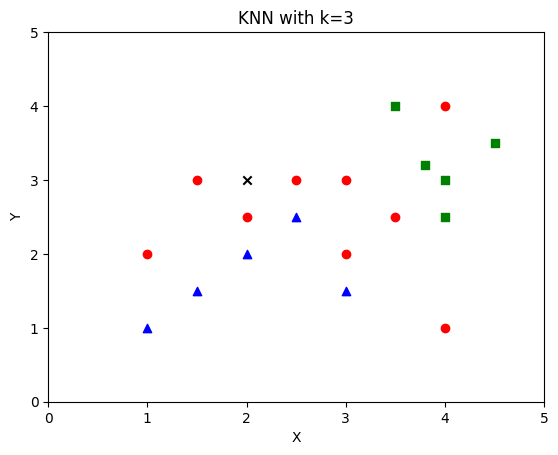
Αξίζει να σημειωθεί ότι υπάρχει η δυνατότητα να προκύψουν Κανόνες Συσχέτισης από ένα Δένδρο Απόφασης. Ειδικότερα, κάθε μονοπάτι από τη ρίζα προς τα φύλλα του δένδρου μπορεί να εκφραστεί ως ένας κανόνας συσχέτισης(Εισαγωγή στην Εξόρυξη Δεδομένων,Tan,Steinbach,Karpatne,Kumar).Η Απόφαση(if[Συνθήκη], then[Απόφαση]) προκύπτει από την ετικέτα της κλάσης σε κάποιο από τα φύλλα του δένδρου.

* **K-Πλησιέστεροι Γείτονες(KNN)**

Η μέθοδος KNN αποτελεί έναν αλγόριθμο κατηγοριοποίησης που λειτουργεί βάσει των k πλησιέστερων γειτόνων στο χώρο των χαρακτηριστικών. Αρχικά το μοντέλο αποθηκεύει και επεξεργάζεται τα δεδομένα εκπαίδευσης. Στη συνέχεια, για να κατηγοριοποιήσει ένα συγκεκριμένο στιγμιότυπο υπολογίζει τις αποστάσεις από όλα τα δεδομένα εκπαίδευσης(όλων των κατηγοριών/κλάσεων). Οπότε, κατηγοριοποιεί το συγκεκριμένο στιγμιότυπο στην κλάση της οποίας τα στιγμιότυπα απέχουν από το προς κατηγοριοποίηση στιγμιότυπο τη μικρότερη απόσταση. Άρα, για κάποιο άγνωστο στιγμιότυπο x πρέπει να βρούμε τα k κοντινότερα δείγματα του συνόλου και να το κατατάξουμε στην κλάση με τα περισσότερα δείγματα μεταξύ των k.

Συμπερασματικά, το στιγμιότυπο κατηγοριοποιείται στην κατηγορία που πλειοψηφεί. Στην περίπτωση που υπάρχει ισοψηφία μεταξύ 2 κλάσεων(k=4, με 2 στοιχεία από την μπλε και 2 από την κόκκινη κατηγορία να είναι τα κοντινότερα εντός της συγκεκριμένης ακτίνας), τότε η επιλογή κατηγορίας γίνεται με τυχαίο τρόπο.

Ακολουθεί ένα παράδειγμα του υλοποιήσαμε σε Python:



Παρατηρούμε ότι το προς κατηγοριοποίηση στιγμιότυπο(x με μαύρο χρώμα) κατηγοριοποιείται στην κατηγορία με τα κόκκινα κυκλάκια, καθώς και τα 3 κοντινότερα στοιχεία(k=3) ανήκουν στη συγκεκριμένη κατηγορία.

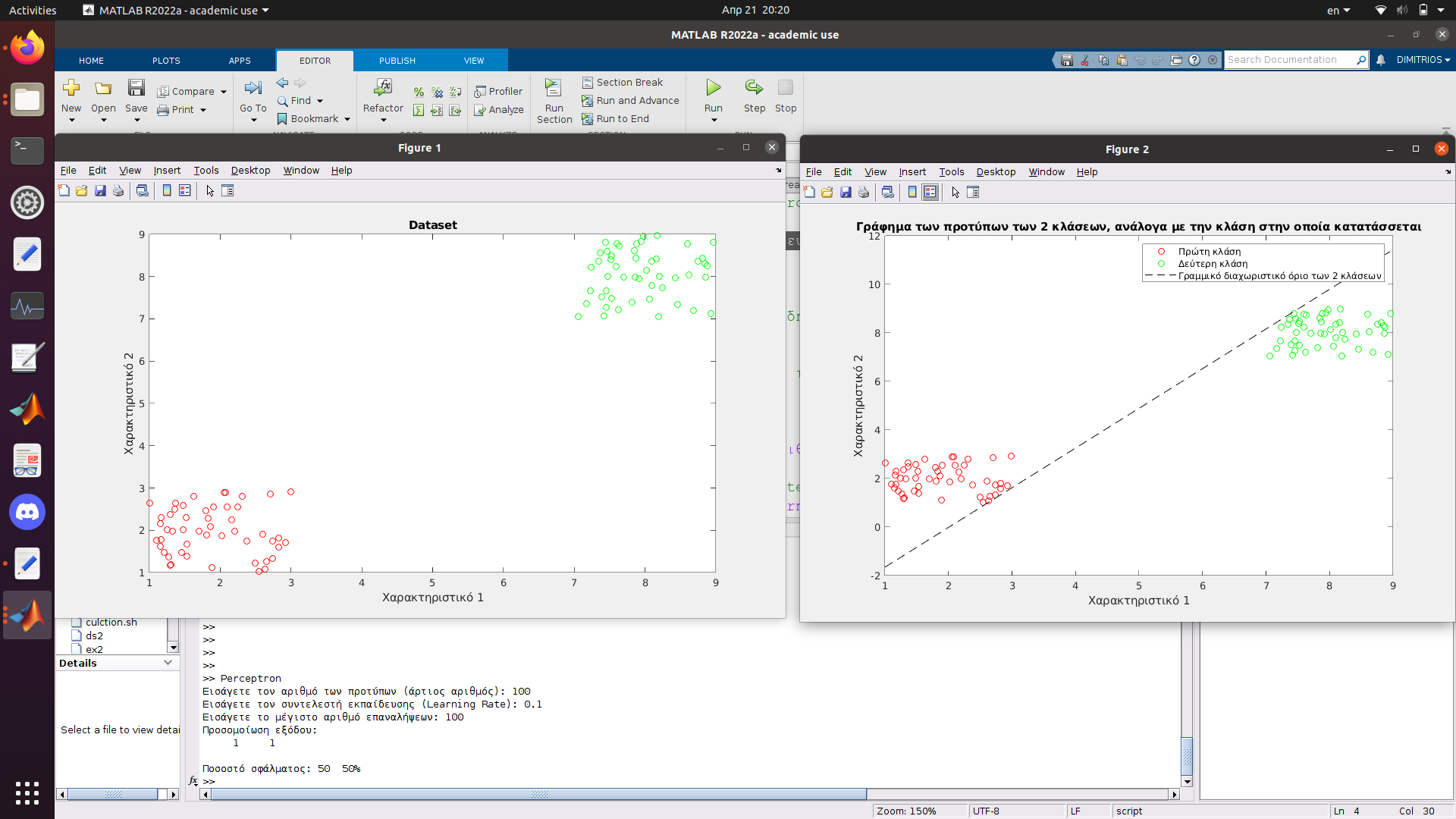
Σε αυτό το σημείο αξίζει να τονιστεί η σημασία επιλογής κατάλληλου k, ανάλογα το πρόβλημα. Αναλυτικότερα, μικρό k οδηγεί σε αύξηση του σφάλματος(υπερπροσαρμογή στα δεδομένα εκπαίδευσης λόγω θορύβου) και μεγάλο k οδηγεί στη συμμετοχή γειτονικών και μη δειγμάτων στην κατηγοριοποίηση. Επίσης, απαιτείται η εύρεση κατάλληλης μετρικής απόστασης, για να επιλέξουμε σωστά τους πλησιέστερους στο στιγμιότυπο γείτονες.

* **Νευρωνικά Δίκτυα**

Καταρχάς, τα νευρωνικά δίκτυα αποτελούν μία υποκατηγορία της Μηχανικής Μάθησης, αλλά και ένα αυτόνομο και αρκετά δημοφιλές ερευνητικό πεδίο. Αναμφίβολα, μιμούνται τον τρόπο με τον οποίο λειτουργεί ο ανθρώπινος εγκέφαλος. Αποτελούνται από νευρώνια(πραγματοποιούν διάφορους υπολογισμούς) και συνάψεις(εκεί αποθηκεύεται η γνώση). Είναι ένας αποδοτικός κατηγοριοποιητής που λειτουργεί με γραμμικά διαχωρίσιμα δεδομένα(διαφάνειες καθηγητή Βασιλά στο μάθημα Νευρωνικά Δίκτυα). Σημαντικό ρόλο διαδραματίζει η επιλογή κατάλληλων εισόδων, συναπτικών βαρών, συνάρτησης ενεργοποίησης, αλλά και ο υπολογισμός της εξόδου του νευρωνίου. Τα βάρη που βρίσκονται ανάμεσα στους νευρώνες φανερώνουν την ισχύ της μεταξύ τους συναπτικής σύνδεσης. Με στόχο να «ταιριάξει» τις εισόδους με κατάλληλες εξόδους πραγμαποιείται ίσως η σημαντικότερη και δυσκολότερη διαδικασία των Νευρωνικών Δικτύων, δηλαδή η προσαρμογή των συναπτικών βαρών, ώστε το δίκτυο να είναι σε θέση να εκτελέσει την επιθυμητή λειτουργία. Η παραπάνω διαδικασία αποτελεί τον αλγόριθμο εκπαίδευσης του δικτύου. Κατά τη συγκεκριμένη διαδικασία το νευρωνικό δίκτυο προσπαθεί να μάθει τα χαρακτηριστικά των δεδομένων και τις αντίστοιχες ετικέτες. Τα νευρωνικά δίκτυα εφαρμόζονται σε αρκετά μεγάλο φάσμα εφαρμογών. Επίσης, υπάρχουν και Πολυστρωματικά Νευρωνικά Δίκτυα, και αν περιέχουν μεγάλο αριθμό κρυφών δικτύων ονομάζονται δίκτυα Βαθιάς Μάθησης. Η Βαθιά Μάθηση έχει παρεισφρήσει και στον τομέα της Εκπαίδευσης, καθώς σχετίζεται με τα Big Data που παρέχουν σχολεία και πανεπιστήμια.

Σε αυτό το σημείο θα περιγράψουμε συνοπτικά το δίκτυο Perceptron, ένα απλό και ιδιαίτερα δημοφιλές νευρωνικό δίκτυο. Περιέχει κόμβους εισόδου και κόμβους εξόδου. Του δίνεται ένα σύνολο εκπαίδευσης, και προσπαθεί με τη βοήθεια του αλγορίθμου εκπαίδευσης που χρησιμοποιεί να ρυθμίσει τα βάρη, ώστε οι πραγματικές έξοδοι του δικτύου(y) να συμβαδίζουν με τις επιθυμητές(d) και να ελαχιστοποιείται το σφάλμα ταξινόμησης((d-y)2). Οπότε, σε κάθε επανάληψη τα βάρη διαφοροποιούνται μέχρι να βρεθούν τα κατάλληλα. Αξίζει να τονιστεί ότι υπάρχουν αρκετά διαχωριστικά όρια που χωρίζουν τα στιγμιότυπα 2 διαφορετικών κατηγοριών. Το Perceptron βρίσκει ένα από αυτά με τη βοήθεια των τιμών των παραμέτρων του. Ψάχνει το υπερεπίπεδο που μηδενίζει τις λάθος ταξινομήσεις και όχι το βέλτιστο.

Ακολουθεί φωτογραφία που οπτικοποιεί τη δράση του Perceptron που περιγράψαμε παραπάνω(εργαστηριακή άσκηση στο περιβάλλον του Matlab για το μάθημα των Νευρωνικών Δικτύων):

****

* **Συσταδοποίηση**

Η Συσταδοποίηση αποτελεί μία αρκετά δημοφιλή μέθοδο Εξόρυξης Δεδομένων που χρησιμοποιείται σε αρκετούς τομείς, οπότε και σε δεδομένα που προέρχονται από τον τομέα της εκπαίδευσης. Είναι η διαδικασία ανάθεσης των σημείων ενός συνόλου σε ομάδες σύμφωνα με μία μετρική απόστασης(Ευκλείδεια απόσταση, Manhattan, απόσταση συνημιτόνου και πολλές άλλες). Ο στόχος της είναι τα στοιχεία της ίδιας ομάδας να έχουν μεγάλη ομοιότητα μεταξύ τους(μικρή απόσταση), ενώ τα σημεία διαφορετικών ομάδων να εμφανίζουν μικρή ομοιότητα (Εξόρυξη από Μεγάλα Σύνολα Δεδομένων,Leskovec, Rajaraman, Ullman). Επιπροσθέτως, χρησιμοποιείται και για την ανακάλυψη κρυφών προτύπων στα δεδομένα. Στην εποχή μας τα προβλήματα συσταδοποίησης είναι αρκετά σύνθετα, καθώς εφαρμόζονται σε σύνολα δεδομένων με αρκετά μεγάλο αριθμό διαστάσεων. Οφείλουμε να σημειώσουμε ότι κάθε συστάδα/ομάδα αντιπροσωπεύεται από ο κέντρο της, δηλαδή το μέσο όρο των σημείων της συστάδας.

Οι αλγόριθμοι συσταδοποίησης χωρίζονται σε 2 μεγάλες κατηγορίες: Παραδοσιακοί και Μοντέρνοι. Υποκατηγορία των παραδοσιακών είναι οι διαμεριστικοί και οι ιεραρχικοί. Ενώ, των μοντέρνων οι πυκνότητας και οι αλγόριθμοι πλέγματος. Αναλυτικότερα, η ιεραρχική συσταδοποίηση χρησιμοποιείται κυρίως για μικρά σύνολα δεδομένων(και για αυτό δε θα αναλυθεί λεπτομερώς). Αρχικά, κάθε στιγμιότυπο αποτελεί τη δική του ομάδα, και συνδυάζουμε μικρότερες ομάδες που βρίσκονται κοντά βάσει απόστασης, ώστε να δημιουργήσουμε μεγαλύτερες, ενώ στο τέλος καταλήγουμε σε μία μόνο ομάδα. Αναπαρίσταται μέσω του δενδρογράμματος.

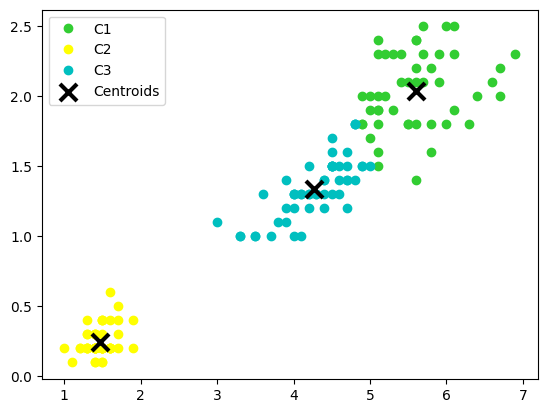
Όσον αφορά τους διαμεριστικούς αλγόριθμους, ο δημοφιλέστερος είναι ο k-means, που είναι ταυτόχρονα απλός και γρήγορος από άποψη εφαρμογής. Η παράμετρος k συμβολίζει των αριθμό των συστάδων. Κάθε συστάδα αντιπροσωπεύεται από το κέντρο της. Για να βρεθεί η βέλτιστη λύση πρέπει να εξεταστούν όλες οι περιπτώσεις. Οπότε, καταλαβαίνουμε ότι αποτελεί μία επαναληπτική διαδικασία όπου εξετάζουμε όλα τα στιγμιότυπα και τα ομαδοποιούμε στην πλησιέστερη προς αυτά ομάδα, με τη χρήση μιας μετρικής απόστασης. Ωστόσο, έχει ευαισθησία στα ακραία δεδομένα και ενδέχεται να συγκλίνει σε τοπικό ελάχιστο. Αρχικά, επιλέγουμε k σημεία ως αρχικά κέντρα βάρους(τυχαία επιλογή). Έπειτα, και μέχρι να μη μεταβάλλονται πλέον τα κέντρα, σχηματίζουμε τις συστάδες και μετακινούμε το κέντρο κάθε συστάδας στο μέσο της. Καθοριστικής σημασίας για τον αλγόριθμο είναι η επιλογή του k, δηλαδή του αριθμού των ομάδων. Κατά την επιλογή του k λαμβάνεται σοβαρά υπόψιν ο αριθμός που οδηγεί σε μικρή τιμή του SSE, όπου SSE το Συνολικό Τετραγωνικό Σφάλμα(συνολικό άθροισμα των τετραγώνων των αποστάσεων μεταξύ κάθε δείγματος και του κέντρου της συστάδας στην οποία ανήκει το δείγμα). Επίσης, ως μέτρο ποιότητας της συσταδοποίησης λογίζεται και ο Συντελεστής Περιγράμματος(Silhouette Coefficient). Αποτελεί ένα μέτρο που χρησιμοποιείται για να παρέχει πληροφορίες για την τοποθέτηση ενός δείγματος στο χώρο, σε σχέση με τις συστάδες. Μεγάλες τιμές του συντελεστή σημαίνουν καλή συσταδοποίηση. Άρα, κατά την επιλογή του k υπολογίζεται ο Συντελεστής Περιγράμματος για διάφορες τιμές του k, και επιλέγεται το k που οδηγεί σε μεγάλη τιμή Συντελεστή Περιγράμματος(κοντά στο 1), εάν αποτελεί το μοναδικό μέτρο. Σε αυτό το σημείο πρέπει να αναφέρουμε ότι σε πραγματικές εφαρμογές δε χρησιμοποιείται μόνο ένα μέτρο υπολογισμού της ποιότητας συσταδοποίησης, αλλά συνδυασμός τους και ανάλυση των αποτελεσμάτων του. Συμπερασματικά, χρειάζεται αρκετή σκέψη για να αποφασιστεί ποιος είναι ο κατάλληλος αριθμός συστάδων για το εκάστοτε παράδειγμα, καθώς οι μικρές τιμές του SSE και τιμές του Συντελεστή Περιγράμματος που πλησιάζουν το 1, είναι συνώνυμες της καλής συσταδοποίησης.

Από την άλλη πλευρά, ο DBSCAN είναι ο δημοφιλέστερος αλγόριθμος πυκνότητας.O DBSCAN είναι κατάλληλος για ομάδες που έχουν υψηλή πυκνότητα σημείων, οι οποίες είναι διαχωρισμένες από άλλα σημεία χαμηλότερης πυκνότητας. Με τον όρο πυκνότητα εννοούμε τον αριθμό σημείων(min\_samples ή MinPts) σε ακτίνα eps(ή Eps). Αρχικά, ο αλγόριθμος ορίζει τα σημεία του συνόλου δεδομένων σε πυρήνα(έχουν πυκνότητα μεγαλύτερη από min\_samples), οριακά έχουν πυκνότητα μικρότερη από min\_samples, αλλά απέχουν από ένα σημείο πυρήνα απόσταση μικρότερη από eps), και θορύβου(τα υπόλοιπα σημεία). Έπειτα, εξαλείφει τα σημεία θορύβου. Επίσης, κάθε ομάδα συνδεδεμένων σημείων πυρήνα αποτελεί χωριστή συστάδα, στην οποία εκχωρούνται τα σημεία ορίου που συμφωνούν με τον παραπάνω περιορισμό. Γενικότερα, ο συγκεκριμένος αλγόριθμος δεν επηρεάζεται από το θόρυβο και χειρίζεται δεδομένα με διαφορετικά σχήματα και μεγέθη. Αντίθετα, εμφανίζει μεγαλύτερη πολυπλοκότητα από τον αλγόριθμο k-means.

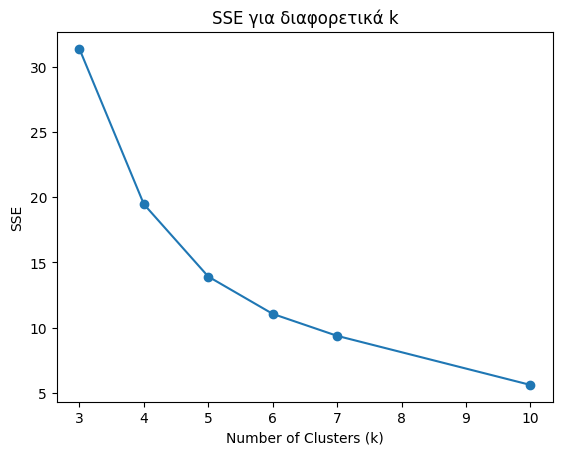
Η επιλογή του τύπου αλγορίθμου συσταδοποίησης που θα επιλέξουμε είναι ανάλογη των δεδομένων στα οποία θέλουμε να εφαρμόσουμε την εκάστοτε μέθοδος Εξόρυξης Δεδομένων.

Ακολουθούν φωτογραφίες που παρουσιάζουν την εφαρμογή των k-means και DBSCAN αντίστοιχα σε δύο σύνολα δεδομένων με αρκετά διαφορετικά χαρακτηριστικά(οι φωτογραφίες προέρχονται από Project στο μάθημα Εξόρυξη Δεδομένων που έχει γίνει σε Python).

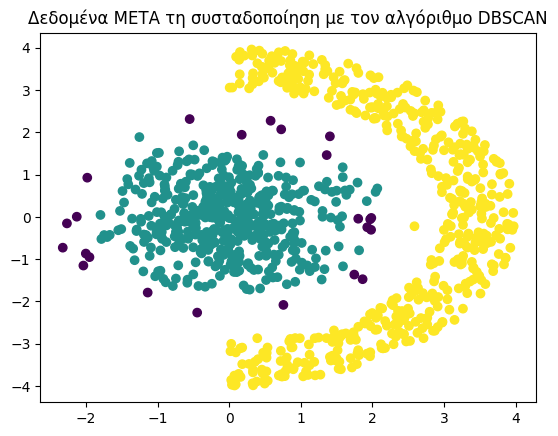
**k-means(με k=3)**



**Επιλογή κατάλληλου k**



**DBSAN(2 κλάσεις: η μία με γαλάζιο, η άλλη με κίτρινα, και τα σημεία θορύβου με σκούρο μπλε)**



* **Πρόβλεψη – Παλινδρόμηση**

Επιλέξαμε να παρουσιάσουμε τις δύο παραπάνω έννοιες ως μία, καθώς είναι άρρηκτα συνδεδεμένες. Αναλυτικότερα, η πρόβλεψη είναι γενικότερος όρος από την παλινδρόμηση, ωστόσο και οι δύο ασχολούνται με την πρόβλεψη μελλοντικών τιμών/γεγονότων. Με τη βοήθεια ενός μοντέλου και χρησιμοποιώντας ορισμένες παρατηρήσεις, είμαστε σε θέση να προβλέψουμε την τιμή κάποιας μεταβλητής(η μεταβλητή μπορεί να είναι τόσο διακριτή, όσο και συνεχής). Αξίζει να σημειωθεί πως κάποια μοντέλα προβλέπουν μόνο την τιμή μόνο μίας μεταβλητής, ενώ άλλα μοντέλα μπορούν να προβλέψουν πολλές μεταβλητές ταυτόχρονα. Είναι απαραίτητη η ύπαρξη ετικετών για τη μεταβλητή εξόδου, που παρέχουν συγκεκριμένες πληροφορίες σχετικά με την τιμή της μεταβλητής εξόδου(Baker,2008). Η πρόβλεψη μπορεί να πραγματοποιηθεί με τη χρήση πολλών τεχνικών, όπως η παλινδρόμηση και τα νευρωνικά δίκτυα. Χωρίς αμφιβολία, η ποιότητα και η αξιοπιστία των προβλέψεων βασίζεται στην ποιότητα των δεδομένων που παρέχονται στο σύστημα.

Η παλινδρόμηση αποτελεί την πιο δημοφιλή τεχνική πρόβλεψης(Chiramdasu,Srivastava,Bhattacharya,Reddy,Gadekallu,2021). Χρησιμοποιείται για συνεχείς, αλλά και διακριτές μεταβλητές και αποτελείται τουλάχιστον από έναν παράγοντα πρόγνωσης. Είναι βασισμένη στη σχέση μεταξύ γεγονότων και μεταβλητών που αναμένεται να προβλεφθούν. Ο στόχος της είναι να βρεθεί η σχέση μεταξύ χαρακτηριστικών του προβλήματος και κάποιας συγκεκριμένης μεταβλητής, ώστε να αποκτήσουμε τη δυνατότητα να προβλέψουμε μελλοντικές τιμές της. Όπως θα δούμε στη συνέχεια του Κεφαλαίου, οι προβλέψεις είναι αρκετά σημαντικές στον τομές της εκπαίδευσης, και αποτελούν προϊόν διεξοδικής ανάλυσης.

* **Εξόρυξη Κειμένου**

Η Εξόρυξη Κειμένου αποτελεί μία μέθοδο που χρησιμοποιείται συχνά στην Εκπαίδευση, αλλά και σε άλλους τομείς. Παρατηρήσαμε ότι δεν αναλύεται παρά μόνο σε ελάχιστα papers που αφορούν την Εξόρυξη Δεδομένων στην Εκπαίδευση. Η πιο δημοφιλής τεχνική της είναι η κατηγοριοποίηση κειμένου, εφαρμόζεται ένας αλγόριθμος, ώστε να ταξινομήσει (εκπαιδευτικά) κείμενα σε διάφορες κατηγορίες(ενότητες), διευκολύνοντας το χρήστη(μαθητής) της εφαρμογής στην οποία περιέχεται να έρθει σε επαφή με το περιεχόμενο που τον ενδιαφέρει(Ray,Saeed,2019). Επίσης, μπορεί να χρησιμοποιηθεί για να βρεθούν κοινά σημεία σε μεγάλο αριθμό εγγράφων(αναζήτηση απάτης), αλλά και να ανακαλυφθούν χρήσιμες και προηγουμένως άγνωστες πληροφορίες.

* **Ανάλυση Κοινωνικών Δικτύων**

Ως τελευταία μέθοδο Εξόρυξης Δεδομένων θα παρουσιάσουμε την Ανάλυση Κοινωνικών Δικτύων, δηλαδή την κατανόηση των σχέσεων μεταξύ των οντοτήτων που βρίσκονται σε μία ομάδα/επιχείρηση/διαδικτυακό lobby. Αν θέλουμε να το εκφράσουμε σχηματικά, τότε οι οντότητες(άνθρωποι) θεωρούνται ως κόμβοι ενός γράφου και οι σχέσεις που έχουν αναπτύξει ως ακμές(Ray,Saeed,2019). Στον τομέα των επιχειρήσεων και της Εκπαίδευσης χρησιμεύει ώστε να αναλυθεί το ποσοστό βοήθειας του κάθε μέλους μιας ομάδας σε ένα project. Επίσης, συνεργάζεται με τη Ψυχολογία(στην Εισαγωγή της εργασίας τονίσαμε τη σημασία της στο χώρο της Εκπαίδευσης), με σκοπό να βοηθήσει συγκεκριμένες οντότητες του δικτύου που χρειάζονται βοήθεια. Τέλος, χρησιμοποιείται για να παράξει ενδιαφέρουσες πληροφορίες που σχετίζονται με την αλληλεπίδραση των οντοτήτων του δικτύου με κάποια εφαρμογή ή εργαλείο που χρησιμοποιούν(για παράδειγμα το eclass ή το Moodle στα ελληνικά πανεπιστήμια), και το πόσο τους βοηθάει να εργαστούν άνετα ή τι προβλήματα τους προκαλεί.

* **Με ποιον τρόπο εφαρμόζονται στην πράξη στον τομέα της Εκπαίδευσης οι παραπάνω μέθοδοι Εξόρυξης Δεδομένων**

1. **Πρόβλεψη της απόδοσης των μαθητών/φοιτητών**

Η συγκεκριμένη διαδικασία, σύμφωνα με τα σχετικά με τον τομέα papers που αναγνώσαμε, αποτελεί την πιο συχνή χρήση των μεθόδων Εξόρυξης Δεδομένων στην Εκπαίδευση(Goyal,Vohra,2012). Χρησιμοποιείται δηλαδή η μέθοδος της παλινδρόμησης ή κάποια τεχνική κατηγοριοποίησης, με στόχο την πρόβλεψη της πορείας ενός μαθητή κατά τη διάρκεια της σχολικής χρονιάς, την πρόβλεψη της βαθμολογίας ενός μαθητή/φοιτητή σε ένα διαγώνισμα, την ικανοποίηση των φοιτητών για τις διαλέξεις ενός συγκεκριμένου μαθήματος, πόσες ώρες διαβάσματος κάθε εβδομάδα είναι κατάλληλες για να περαστεί ένα μάθημα με επιτυχία στο πανεπιστήμιο ή το πόσο πιθανό είναι ένα εργαλείο e-learning να βοηθήσει τους μαθητές. Γίνεται δηλαδή προσπάθεια πρόβλεψης της τιμής κάποιας μεταβλητής. Εάν η τιμή είναι συνεχής χρησιμοποιούμε παλινδρόμηση, ενώ εάν είναι διακριτή επιλέγουμε κατηγοριοποίηση (Goyal,Vohra,2012). Οι τεχνικές κατηγοριοποίησης που χρησιμοποιούνται για τη συγκεκριμένη λειτουργία είναι κυρίως τα νευρωνικά δίκτυα και οι κανόνες συσχέτισης.

1. **Πρόβλεψη πιθανής εγκατάλειψης στην τριτοβάθμια εκπαίδευση**

Ένα συχνό φαινόμενο τόσο στην Ελλάδα, όσο και σε άλλες ευρωπαϊκές χώρες είναι η εγκατάλειψη ενός συγκεκριμένου μαθήματος στη μέση του εξαμήνου ή ακόμη σοβαρότερο είναι το ζήτημα της γενικότερης εγκατάλειψης των σπουδών για μαθησιακούς λόγους. Τα παραπάνω ζητήματα είναι πολυσύνθετα και απαιτείται η τροποποίηση όλου του συστήματος εκπαίδευσης, ωστόσο η Εξόρυξη Δεδομένων βοηθάει μόνο ως διορθωτικό μέτρο. Αρχικά, απαιτείται η συλλογή αρκετών δεδομένων που αφορούν κάθε εξάμηνο της φοίτησης των μαθητών(βαθμολογίες, αριθμός περασμένων μαθημάτων, συμμετοχή στις διαλέξεις, χρόνος απασχόλησης με πλατφόρμες e-learning που παρέχει το πανεπιστήμιο, ενδιαφέρον του φοιτητή για το συγκεκριμένο επιστημονικό τομέα και πολλά άλλα). Ακολουθεί η προεπεξεργασία και ανάλυση των δεδομένων. Επίσης, κρίνεται ως απαραίτητη και η ανάλυση ακραίων τιμών(outliers). Τα outliers αποτελούν στιγμιότυπα του συνόλου δεδομένων που παρουσιάζουν διαφορετική «συμπεριφορά» από τα υπόλοιπα, δηλαδή δεν ακολουθούν τη γενική κατανομή. Στην τριτοβάθμια εκπαίδευση είναι πιθανό ορισμένα outliers αν δεν ανακαλυφθούν έγκαιρα να αποτελέσουν φοιτητές που εγκαταλείπουν τις σπουδές τους χωρίς να τις ολοκληρώσουν. Για την επίτευξη και της συγκεκριμένης πρόβλεψης χρησιμοποιούνται διάφοροι αλγόριθμοι κατηγοριοποίησης(Dekker,2009),αλλά κυρίως τα νευρωνικά δίκτυα.

1. **Διαφοροποίηση της μορφής των διαγωνισμάτων – Τροποποίηση του Προγράμματος Σπουδών**

Αναμφίβολα υπάρχουν αρκετά παράπονα μαθητών και φοιτητών για τον τρόπο με τον οποίο πραγματοποιούνται οι εξετάσεις τόσο στα σχολεία όσο και στα πανεπιστήμια. Για παράδειγμα, μέχρι και σήμερα εμφανίζεται το φαινόμενο της συγγραφής κώδικα σε χαρτί σε εξετάσεις σχολών Πληροφορικής ελληνικών πανεπιστημίων. Ενώ, σε όλα τα επίπεδα της εκπαίδευσης έχει καθιερωθεί η απομνημόνευση χωρίς να επιτυγχάνεται η πραγματική κατανόηση. Σε πολλές περιπτώσεις φαίνεται πως έχει επικρατήσει και η βαθμοθηρία. Από την άλλη πλευρά, στα πανεπιστήμια είναι σύνηθες το γεγονός εμφάνισης των ίδιων εννοιών με τον ίδιο ακριβώς τρόπο σε αρκετά μαθήματα. Επίσης, η παρακολούθηση κάποιων μαθημάτων απαιτεί την επιτυχή παρακολούθηση κάποιου άλλου μαθήματος στο παρελθόν, που ωστόσο δεν το έχουν δηλώσει ακόμη όλοι οι φοιτητές του συγκεκριμένου εξαμήνου. Συχνό είναι και το φαινόμενο αρκετοί μαθητές/φοιτητές να ξεχάσουν έννοιες που έχουν διδαχθεί στο παρελθόν. Οπότε, συμπεραίνουμε ότι και το Πρόγραμμα Σπουδών στα σχολεία και τα πανεπιστήμια πρέπει να αναδιαμορφωθεί. Για τους παραπάνω λόγους και αρκετούς ακόμη(που σίγουρα θα υπάρχουν), πρέπει με τη βοήθεια των κανόνων συσχέτισης αλλά και κάποιων άλλων μεθόδων ταξινόμησης να δημιουργούνται εξατομικευμένα διαγωνίσματα, που θα έχουν ως αποτέλεσμα την επίλυση συγκεκριμένων αποριών του εκάστοτε μαθητή(Romero,2013). Ακόμη, με την είσοδο της Εξόρυξης Δεδομένων στην εκπαίδευση είναι δυνατή και η real-time αξιολόγηση των μαθητών/φοιτητών μέσω της χρήσης εργαλείων e-learning(Ray,Saeed,2018). Επιπροσθέτως η παραπάνω τακτική μπορεί να βοηθήσει και τους καθηγητές να προσαρμόσουν το μάθημα τους, όπως και το Υπουργείο Παιδείας να ακολουθήσει την ίδια τακτική για το Πρόγραμμα Σπουδών. Με τη χρήση των κανόνων συσχέτισης είναι δυνατή η αναγνώριση προτύπων που αφορούν τη συμπεριφορά των μαθητών κατά τη διάρκεια μιας διάλεξης και την απόδοση τους στα διαγωνίσματα. Με αυτό τον τρόπο ο καθηγητής μπορεί να αναγνωρίσει τις ανάγκες των μαθητών του, να αντιληφθεί ποιες μέθοδοι μάθησης τους βοηθούν πραγματικά και να προσαρμόσει ανάλογα το μάθημα. Όσον αφορά το Πρόγραμμα Σπουδών, υπάρχει η δυνατότητα αλλαγής της δομής του, με στόχο να συμβαδίζει με τις ανάγκες της εποχής μας και να βοηθάει μαθητές και καθηγητές να αποδίδουν καλύτερα και να είναι χαρούμενοι στο χώρο του σχολείου ή του πανεπιστημίου. Για την τροποποίηση του Προγράμματος Σπουδών χρησιμοποιούνται επίσης οι κανόνες συσχέτισης, και τα δένδρα απόφασης(Wu,Liao,Hsia,Tennyson,2008).

1. **Ομαδοποίηση μαθητών**

Όπως τονίσαμε και σε προηγούμενο σημείο του Κεφαλαίου, η μέθοδος της ομαδοποίησης είναι μία από τις σημαντικότερες τεχνικές Εξόρυξης Δεδομένων. Στον τομέα της Εκπαίδευσης χρησιμοποιείται για την ταξινόμηση μαθητών/φοιτητών σε ομάδες με βάση κάποια κοινά χαρακτηριστικά τους. Επιλέγονται κάποια κριτήρια όπως παρόμοια μαθησιακά χαρακτηριστικά, παρεμφερείς βαθμολογίες σε διαγωνίσματα του συγκεκριμένου μαθήματος ή παρόμοιοι βαθμοί σε προαπαιτούμενα μαθήματα περασμένων ετών. Στόχος είναι να δημιουργηθούν ομάδες με υψηλή ομοιότητα, για να μπορέσει ο καθηγητής να τους παρέχει ένα εξατομικευμένο σύστημα μάθησης. Οι μορφές ομαδοποίησης που χρησιμοποιούνται στο συγκεκριμένο τομέα είναι η ιεραρχική ομαδοποίηση και ο k-means. Καθοριστικός θα είναι και ο ρόλος του εργαλείου e-learning που χρησιμοποιείται για να μπορεί ο καθηγητής να παρακολουθεί την πορεία του κάθε μαθητή, αλλά και των ομάδων, μέσω της ανάλυσης των δεδομένων που παρέχουν οι πλατφόρμες e-learning. Σε αυτό το σημείο αξίζει να σημειώσουμε ότι είναι απαραίτητη η διαχείριση της όλης διαδικασίας εκ μέρους του καθηγητή, ώστε όλοι οι μαθητές/φοιτητές να καταλάβουν τη σημασία της ομαδοποίησης και να μη τη συνδέσουν ατυχώς με περιθωριοποίηση ή να μη τους επηρεάσει αρνητικά στην ψυχολογία τους. Οπότε, καταλαβαίνουμε ότι η συγκεκριμένη μέθοδος παρουσιάζει αυξημένη δυσκολία λόγω κοινωνικών και ψυχολογικών επιδράσεων που μπορεί να αποφέρει τόσο στον μαθητή όσο και στην τάξη γενικότερα.

1. **Πρόβλεψη των μαθημάτων που δηλώνουν οι φοιτητές**

Η διαδικασία επιλογής μαθημάτων(αναφερόμαστε σε μαθήματα επιλογής και όχι υποχρεωτικά) και κυρίως η σειρά με την οποία θα δηλωθούν είναι μία διαδικασία που «ταλαιπωρεί» τους φοιτητές των μεγαλύτερων εξαμήνων. Η συγκεκριμένη διαδικασία εξαρτάται από πολλούς παράγοντες, όπως οι βαθμοί του φοιτητή σε ορισμένα μαθήματα, η φιλοσοφία του καθηγητή που το διδάσκει, ο τρόπος που διδάσκεται και εξετάζεται το μάθημα και το ενδιαφέρον του φοιτητή για το επιστημονικό πεδίο που ανήκει το μάθημα. Με τη βοήθεια νευρωνικών δικτύων που δέχονται ως είσοδο τα παραπάνω δεδομένα(Guo,2010), και αναλογιζόμενοι τα ενδιαφέροντα του εκάστοτε φοιτητή πραγματοποιείται πρόβλεψη των μαθημάτων επιλογής που είναι κατάλληλα για τον ίδιο.

1. **Ανίχνευση των σχέσεων μεταξύ μαθητών/φοιτητών**

Στην Εισαγωγή της εργασίας μας αναφέραμε ότι ο τομέας της Εκπαίδευσης δεν αποτελείται αποκλειστικά από τη διαδικασία της μάθησης, αλλά ασχολείται και με θέματα αξιών, κοινωνικοποίησης και κανόνες συμπεριφοράς. Αρκετά σημαντικές είναι και οι σχέσεις που δημιουργούνται μεταξύ των μαθητών/φοιτητών, αλλά και των καθηγητών με τους μαθητές/φοιτητές. Αναμφίβολα, για να κριθεί επιτυχημένη η διαδικασία της μάθησης πρέπει να υπάρχει καλό κλίμα στο χώρο των σχολείων και των πανεπιστημίων. Όσον αφορά τις μεθόδους Εξόρυξης Δεδομένων που χρησιμοποιούνται για την επίτευξη των παραπάνω, η Ανάλυση Κοινωνικών Δικτύων δύναται να εντοπίσει τις σχέσεις των οντοτήτων που βρίσκονται σε μία τάξη. Μπορούμε δηλαδή να εξάγουμε συμπεράσματα για τον τρόπο που επηρεάζουν την απόδοση των μαθητών αλλά και γενικότερα την εικόνα τις τάξεις οι συγκεκριμένες σχέσεις. Επίσης, η ΑΚΔ είναι απαραίτητη για τους καθηγητές, για να καταλάβουν τη συνεισφορά των μελών μιας ομάδας σε ομαδικές εργασίες και γενικότερα για τι πραγματικά συμβαίνει στην τάξη.

1. **Ανάλυση του περιεχομένου του μαθησιακού υλικού**

Πιθανώς όλοι οι φοιτητές έχουν παρακολουθήσει μαθήματα των οποίων το eclass(ή άλλη αντίστοιχη πλατφόρμα) είναι ιδιαίτερα πολύπλοκο, δηλαδή περιέχει πολλά αρχεία τα οποία είναι μπερδεμένα, και έχουν δημιουργηθεί διαφορετική χρονική περίοδο. Αναντίρρητα υπάρχουν αρκετά παρόμοια θέματα που τόσο μαθητές όσο και φοιτητές έχουν έρθει αντιμέτωποι. Η Εξόρυξη Κειμένου παρέχει τη σύνοψη του περιεχομένου του μαθησιακού υλικού που φιλοξενείται σε πλατφόρμες e-learning. Επιπλέον, βοηθάει στην εύρεση λέξεων-κλειδιών και γενικά σημαντικών εννοιών που εμφανίζονται σε αρχεία που περιέχουν σημειώσεις μαθημάτων και το υλικό των δια ζώσης ή των εξ αποστάσεως διαλέξεων, βοηθώντας τους μαθητές/φοιτητές, αλλά και τους καθηγητές.

***Κεφάλαιο 3: Χρήση διαφόρων εργαλείων και τεχνικών για τη συλλογή, την αποθήκευση και την επεξεργασία εκπαιδευτικών δεδομένων***

Αναντίρρητα, τα δεδομένα αποτελούν το κεντρικό σημείο οποιασδήποτε διαδικασίας ανάλυσης δεδομένων. Κατανοούμε δηλαδή τη σημασία τους για όλα τα βήματα της διαδικασίας. Η συλλογή τους αποτελεί την απόκτηση δομημένων ή αδόμητων πληροφοριών, οι οποίες είναι απαραίτητες για την εξόρυξη γνώσης. Ειδικά στον τομέα με τον οποίο ασχολούμαστε, την Εκπαίδευση, η συλλογή δεδομένων είναι αρκετά σημαντική και δύσκολη διαδικασία. Πρέπει να γίνεται προσεκτικά και να συμβαδίζει με τους σχετικούς νόμους που έχουν θεσπιστεί. Είναι κατανοητό το γεγονός ότι στα σύνολα δεδομένων που δημιουργούνται στον εκπαιδευτικό τομέα υπάρχουν αρκετά ευαίσθητα δεδομένα, και αφού παράγονται από παιδιά και εφήβους πρέπει να αποθηκεύονται σε προστατευμένους χώρους που δεν επιτρέπουν την ανάγνωση ή την επεξεργασία τους σε μη διαπιστευμένα άτομα. Στις μέρες μας τα δεδομένα εκπαίδευσης, αλλά και γενικότερα τα δεδομένα οποιουδήποτε τομέα, φιλοξενούνται σε σύγχρονα μηχανήματα με μεγάλο αποθηκευτικό χώρο, ώστε να μπορούν να εξυπηρετήσουν σύνολα δεδομένων που αποτελούν Big Data. Η επεξεργασία δεδομένων αναφέρεται στην παραγωγή αποτελεσμάτων από τα δεδομένα, μέσω της εφαρμογής διάφορων μεθόδων. Να σημειώσουμε ότι σημαντική είναι και η διαδικασία που προηγείται της επεξεργασίας, δηλαδή η προεπεξεργασία, καθώς αφορά την προετοιμασία των δεδομένων(καθαρισμός δεδομένων, απαλοιφή θορύβου, μετασχηματισμός, ανάλυση συσχέτισης) πριν την εφαρμογή μεθόδων ανάλυσης δεδομένων.

Καταρχάς, η συλλογή εκπαιδευτικών δεδομένων είναι μία διαδικασία κατά την οποία πραγματοποιείται η λήψη δεδομένων που αφορούν μαθητές/φοιτητές, καθηγητές, τις σχέσεις που αναπτύσσουν οι διάφορες οντότητες μεταξύ τους και γενικότερα τη διδασκαλία. Πραγματοποιείται με στόχο να βρεθούν δεδομένα τα οποία έπειτα από την ανάλυση τους θα βοηθήσουν τα μέλη της εκπαιδευτικής κοινότητας και το Υπουργείο Παιδείας να ανιχνεύσουν λάθη τα οποία γίνονται στην εκπαιδευτική διαδικασία και να τη βελτιώσουν στα σημεία που το έχει ανάγκη. Η γνώση που παράγεται μετά τη διαδικασία της ανάλυσης δεδομένων, μπορεί να χρησιμοποιηθεί για το σχεδιασμό γενικών δράσεων που θα βοηθήσουν όλους τους μαθητές/φοιτητές ή και ατομικών παρεμβάσεων, δηλαδή βοήθειας σε μαθητές/φοιτητές και καθηγητές που το έχουν πραγματικά ανάγκη(σε τέτοιες περιπτώσεις πρέπει οι φορείς εκπαίδευσης να κινηθούν γρήγορα και αποτελεσματικά).

Οι υπεύθυνοι της λήψης αποφάσεων στον τομέα της εκπαίδευσης, συνεργαζόμενοι με ειδικούς από το χώρο της Πληροφορικής, έχουν θεσπίσει συγκεκριμένα εργαλεία και τεχνικές για τη συλλογή εκπαιδευτικών δεδομένων. Αναλυτικότερα, πρέπει να γίνει σαφές το είδος των δεδομένων που συλλέγονται. Αυτό εξαρτάται από τις συνθήκες κάτω από τις οποίες θα πραγματοποιηθεί η συλλογή, αλλά και από τους στόχους της διαδικασίας ανάλυσης δεδομένων που έχουν γίνει γνωστοί πριν από την έναρξη της όλης διαδικασίας. Σε αυτό το σημείο θα αναφέρουμε τους δημοφιλέστερους τύπους εκπαιδευτικών δεδομένων που καταγράφονται κατά τη διαδικασία της συλλογής. Αρχικά, είναι αρκετά σημαντική η καταγραφή των απουσιών μαθητών από το σχολείο και αν υπάρχει ενημέρωση για τους λόγους που πραγματοποιείται. Επίσης, χρειάζονται δεδομένα που σχετίζονται με προσωπικά/δημογραφικά στοιχεία(ηλικία, εθνικότητα, μορφωτικό επίπεδο, …) των μαθητών, των γονέων τους , αλλά και των καθηγητών. Επιπλέον , πρέπει να καταγράφονται βασικά στοιχεία της συμπεριφοράς μαθητών/φοιτητών, αλλά και καθηγητών από τους μαθητές και τους γονείς τους, με σκοπό να προβλεφθούν επικίνδυνες καταστάσεις σε πολλά επίπεδα. Δηλαδή, η εξόρυξη γνώσης δεν περιορίζεται αποκλειστικά στη μάθηση, αλλά καταγράφει και στοιχεία που απαιτείται η συνεργασία με ειδικούς ψυχολόγους και κοινωνιολόγους για την ανάλυση τους. Αναμφίβολα, απαιτείται η καταγραφή στο σύστημα των βαθμών μαθητών και φοιτητών σε εργασίες και εξετάσεις, αλλά και βαθμολογίες προηγούμενων ετών. Ακόμη, χρειάζεται η καταγραφή πληροφοριών που αφορούν τη ζωή των μαθητών(για παράδειγμα αν ασχολούνται με κάποιο άθλημα) και των φοιτητών(για παράδειγμα αν εργάζονται) τις ώρες που δε βρίσκονται στο σχολείο και το πανεπιστήμιο αντίστοιχα.

Από τους τύπους δεδομένων που αναφέραμε προκύπτουν και οι τεχνικές συλλογής δεδομένων που εφαρμόζονται στην εκπαίδευση. Στα περισσότερα papers που αναγνώσαμε, τα ερωτηματολόγια αναφέρονται ως η δημοφιλέστερη τεχνική. Συμπληρώνονται από τους μαθητές/φοιτητές και αναφέρονται τόσο στους ίδιους(σχετικά με την απόδοση τους) όσο και στους καθηγητές τους και τον τρόπο με τον οποίο πραγματοποιείται η διδασκαλία. Αντίστοιχα, και οι καθηγητές συμπληρώνουν παρόμοια ερωτηματολόγια, για τους στόχους και την πορεία του μαθήματος. Τα ερωτηματολόγια(κλειστού και ανοιχτού τύπου) μπορούν να διατεθούν προς συμπλήρωση σε διάφορες χρονικές περιόδους μέσα στη χρονιά. Επίσης, αρκετά συχνή είναι και η συλλογή δεδομένων μέσω συζητήσεων με μαθητές ή φοιτητές σε μικρές ομάδες ή και ατομικά με την παρουσία εκπαιδευτικών φορέων ή ψυχολόγων. Γίνονται συζητήσεις που αφορούν τόσο τη μάθηση όσο και κοινωνικά θέματα. Συνεχίζουμε με την τεχνική της παρατήρησης. Με την καταγραφή πληροφοριών μέσω της παρατήρησης οι ερευνητές κατανοούν το εκπαιδευτικό περιβάλλον χωρίς να χρειάζεται να βρίσκονται σε αυτό. Η παρατήρηση μπορεί να επιτευχθεί μέσω της παρακολούθησης των μαθημάτων από φορείς της εκπαίδευσης, αλλά και με τη βοήθεια των καθηγητών και σημειώσεων που κρατούν κατά τη διάρκεια των μαθημάτων σε όλη τη σχολική χρονιά. Γίνεται με αυτό τον τρόπο γνωστή η εκπαιδευτική τακτική που ακολουθεί ο καθηγητής, εξετάζεται αν γίνεται κατανοητή από τους μαθητές/φοιτητές και τι σχέσεις έχουν αναπτύξει οι τελευταίοι μεταξύ τους, αλλά και με τους καθηγητές τους. Τέλος, δεν πρέπει να παραλείψουμε την καταγραφή πληροφοριών που βρίσκονται σε έγγραφα. Τόσο στα σχολεία όσο και στα πανεπιστήμια υπάρχουν πολυάριθμα έγγραφα, πολλά από τα οποία είναι αρκετά σημαντικά για την εξόρυξη γνώσης, ωστόσο βρίσκονται σε γραφεία και αποθήκες. Με την καταγραφή κάποιων σημαντικών πληροφοριών που περιέχουν τα έγγραφα και αφορούν οποιαδήποτε δραστηριότητα διενεργείται σε ένα εκπαιδευτικό περιβάλλον, επιτρέπουν τη σύγκριση με παλαιότερες εποχές και ίσως αποτρέψουν λάθη τα οποία έγιναν στο παρελθόν.

Συνεχίζοντας με τα εργαλεία συλλογής δεδομένων, χωρίζονται σε δύο κατηγορίες, στην πρώτη κατηγορία ανήκουν αυτά τα οποία συλλέγουν αυτόματα πληροφορίες κατά την αλληλεπίδραση του μαθητή ή του καθηγητή με κάποια εφαρμογή e-learning και στη δεύτερη εκείνα τα οποία συλλέγονται μέσω εγγράφων, συζητήσεων και καταγραφής συμπεριφορών στο χώρο του σχολείου ή του πανεπιστημίου. Η πρώτη κατηγορία αναφέρεται ως ενεργή συλλογή δεδομένων, και πραγματοποιείται σε πραγματικό χρόνο(Gustafsson-Wright, Osborne, Aggarwal,2022). Αξίζει να τονίσουμε ότι διαφορετικά εργαλεία που ανήκουν στη δεύτερη κατηγορία συλλέγουν δεδομένα ανά διαφορετική χρονική συχνότητα, που εξαρτάται από το είδος της τεχνικής που χρησιμοποιείται. Όπως αναφέρουν οι Gustafsson-Wright, Osborne και Aggarwal υπάρχουν εργαλεία που πραγματοποιούν τις διαδικασίες της συλλογής και αποθήκευσης δεδομένων που σχετίζονται με τη συμμετοχή στις εκπαιδευτικές διαδικασίες, τις παρατηρήσεις και τη διαχείριση αξιολογήσεων. Αναφέρουν ότι τα δεδομένα που συλλέγονται με τέτοιου τύπου εργαλεία αναλύονται μέσω της εφαρμογής υπολογιστικών φύλλων. Επίσης, ανακτήσαμε το εργαλείο Waliku. Αποτελεί ένα ψηφιακό εργαλείο που σχετίζεται με την εκπαίδευση. Σχετίζεται κυρίως με τη συλλογή δεδομένων που προέρχονται από έγγραφα και έχει ως στόχο τη διαχείριση της παιδικής ευημερίας. Ακόμη, χρησιμοποιείται και για την αξιολόγηση προγραμμάτων σπουδών. Είναι γνωστό για την παροχή βοηθητικών οπτικοποιήσεων που αφορούν τα αποτελέσματα της ανάλυσης των δεδομένων που συλλέγονται. Στη συνέχεια, θα αναφέρουμε το Chimple(παρέχεται ως εφαρμογή για κινητά τηλέφωνα). Εργαλεία τέτοιου τύπου συλλέγουν δεδομένα και παράγουν εκπαιδευτικό περιεχόμενο, ώστε οι χρήστες του εργαλείου(μαθητές μικρών τάξεων) να αποκτήσουν και γνώσεις κατά την επαφή τους με αυτό. Υπάρχουν και άλλα εργαλεία της ίδιας κατηγορίας, τα οποία με τη μορφή εκπαιδευτικών παιχνιδιών και σε συνεννόηση φυσικά με εκπαιδευτικούς φορείς και το Υπουργείο Παιδείας της εκάστοτε χώρας, συλλέγουν εκπαιδευτικά δεδομένα. Τα εργαλεία που ανήκουν στη συγκεκριμένη κατηγορία είναι ευκόλως κατανοητό ότι συλλέγουν αυτόματα δεδομένα κατά τη διάρκεια της αλληλεπίδρασης του μαθητή με την εφαρμογή. Το τελευταίο εργαλείο που αναφέρεται ονομάζεται Tangerine, και ανήκει στην κατηγορία εργαλείων συλλογής και ανάλυσης δεδομένων. Περιλαμβάνει και όλες τις υπόλοιπες κατηγορίες εργαλείων, καθώς οπτικοποιεί συγκεκριμένα δεδομένα, αλλά παράγει και γνώσεις στους χρήστες του. Συλλέγει αυτόματα και σε πραγματικό χρόνο δεδομένα. Λαμβάνει δεδομένα μέσω αξιολογήσεων μαθητών μετά από κάποιου είδους εξέταση γνώσεων σε συγκεκριμένους τομείς εκπαίδευσης.

Επιπλέον, βρήκαμε και κάποια ακόμη εργαλεία συλλογής εκπαιδευτικών δεδομένων. Το Moodle παρέχει τη δυνατότητα της συλλογής δεδομένων που σχετίζονται με τη συμμετοχή φοιτητών στα μαθήματα, τους βαθμούς τους σε εργασίες και σε εξετάσεις. Οι εφαρμογές Qualtrics και Super Monkey παρέχουν τη δυνατότητα δημιουργίας ερωτηματολογίων και ανάλυσης τους για οποιοδήποτε τομέα, οπότε μπορούν να χρησιμοποιηθούν και για την εκπαίδευση. Τέλος, ένα δημοφιλές εργαλείο παρατήρησης συμπεριφορών μέσα στην τάξη είναι το Classroom Walkthrough και χρησιμοποιείται για τη βελτίωση της διδασκαλίας.

Από τα παραπάνω καταλάβαμε ότι εκτός από τις παραδοσιακές πρακτικές εκπαίδευσης, τις τελευταίες δεκαετίες λόγω της Ψηφιακής Επανάστασης παρατηρείται εισροή νέων εκπαιδευτικών συστημάτων που βασίζονται στην Πληροφορική και ειδικότερα στις τεχνικές Τεχνητής Νοημοσύνης. Οπότε, έχουν δημιουργηθεί και νέες πηγές δεδομένων, κάποιες από τις οποίες ήδη αναφέραμε. Αναμφίβολα, κάθε μία από τις νέες πηγές δεδομένων πρέπει να αποθηκευτεί σε συγκεκριμένο χώρο και να υποβληθεί σε επεξεργασία με διαφορετικές τεχνικές ανάλογα τα χαρακτηριστικά της.

Ο πιο δημοφιλής χώρος φιλοξενίας δεδομένων είναι οι σχεσιακές βάσεις δεδομένων. Αποτελούνται από πίνακες, κάθε ένας από τους πίνακες αποτελείται από διάφορες στήλες στις οποίες υπάρχουν εγγραφές/στιγμιότυπα. Οι πίνακες συνδέονται μέσω συσχετίσεων, που εξαρτώνται από το είδος του προβλήματος. Για να αντλήσουμε δεδομένα από μία σχεσιακή βάση δημιουργούμε ερωτήματα σε γλώσσα SQL. Από τα εργαλεία συλλογής δεδομένων που αναφέραμε παραπάνω, το Moodle διαθέτει σχεσιακή βάση δεδομένων, για να αποθηκεύει στοιχεία φοιτητών. Η τεχνική επεξεργασίας των συγκεκριμένων εκπαιδευτικών δεδομένων ονομάζεται σχεσιακή Εξόρυξη Δεδομένων(Romero,Romero,Ventura,2014) και αποτελεί υποκατηγορία της Εξόρυξης Δεδομένων που μελετήσαμε σε προηγούμενα κεφάλαια της εργασίας μας. Οι μέθοδοι σχεσιακής Εξόρυξης Δεδομένων αναζητούν μοτίβα μεταξύ των διαφορετικών πινάκων της ΒΔ.

Σε αυτό το σημείο θα αναφέραμε ένα δεύτερο είδος βάσεων δεδομένων, τις χρονικές. Αποθηκεύουν δεδομένα που σχετίζονται άμεσα με το χρόνο, δηλαδή ακολουθίες γεγονότων ή τιμών που συλλέγονται ανά τακτά χρονικά διαστήματα. Οι περισσότερες εφαρμογές e-learning δημιουργούν αρχείο καταγραφής του κάθε μαθητή ή φοιτητή. Καταγράφουν δηλαδή δεδομένα που αφορούν το αν ο μαθητής/φοιτητής έχει κάνει απουσίες και πότε, πόσο συχνά επισκέπτεται την εφαρμογή e-learning και για ποια μαθήματα και ενότητες μαθημάτων συγκεκριμένα, πότε υπέβαλλε την εκάστοτε εργασία και γενικότερα κάθε ενέργεια που έχει πραγματοποιήσει. Η τεχνική επεξεργασίας που χρησιμοποιείται για τα παραπάνω δεδομένα είναι και αυτή ένα είδος Εξόρυξης Δεδομένων και ονομάζεται εξόρυξη διαδοχικών προτύπων(Romero,Romero,Ventura,2014).

Στη συνέχεια, οφείλουμε να περιγράψουμε τις βάσεις δεδομένων εγγράφων. Φιλοξενούν έγγραφα από ετερογενείς πηγές όπως εργασίες, βιβλία, σημειώσεις, αξιολογήσεις, και κείμενα που προέρχονται από συνομιλίες μαθητών με καθηγητές ή μεταξύ τους. Καταλαβαίνουμε ότι η επεξεργασία των εγγράφων είναι μια πολύ σημαντική διαδικασία στον τομέα της εκπαίδευσης και αν πραγματοποιούταν με τις παραδοσιακές μεθόδους θα ήταν υπερβολικά χρονοβόρα έως και αδύνατη. Στα έγγραφα βρίσκεται αρκετά σημαντική πληροφορία που πρέπει να μετασχηματιστεί σε γνώση, ώστε να βελτιστοποιηθεί η λήψη αποφάσεων από τη μεριά του Υπουργείου Παιδείας και των εκπαιδευτικών φορέων. Η δημοφιλέστερη τεχνική επεξεργασίας εγγράφων είναι η εξόρυξη κειμένου, μέθοδος που αναλύθηκε σε προηγούμενο κεφάλαιο της παρούσας εργασίας.

Ένα ακόμη είδος βάσης δεδομένων που διαβάσαμε σε σχετικά με το θέμα papers είναι οι ΒΔ πολυμέσων. Τα δεδομένα που αποθηκεύουν είναι κυρίως εικόνες και βίντεο. Τα παραπάνω δεδομένα είναι αρκετά «βαριά» για ένα σύστημα, οπότε πρέπει να γίνεται προσεκτικά και η αποθήκευση τους, αλλά και η επεξεργασία τους. Ως εικόνες μπορούν να θεωρηθούν γραπτές εργασίες μαθητών/φοιτητών σε χαρτί κατά τη διάρκεια διαδικτυακών μαθημάτων και εξετάσεων. Παράλληλα, εικόνες υπάρχουν στο μεγαλύτερο μέρος των σημειώσεων και γενικά των αρχείων που υπάρχουν σε εφαρμογές e-learning. Επίσης, ως βίντεο θεωρούνται διάφορα tutorials που μπορεί να βοηθούν στην τριτοβάθμια εκπαίδευση που υπάρχουν εργαστηριακά μαθήματα και χρησιμεύουν στην εκμάθηση κάποιας εφαρμογής/εργαλείου. Η τεχνική επεξεργασίας που χρησιμοποιείται για δεδομένα που φιλοξενούνται σε βάσεις δεδομένων πολυμέσων είναι η Εξόρυξη Δεδομένων πολυμέσων, και όπως και οι υπόλοιπες τεχνικές επεξεργασίας που αναφέραμε μέχρι αυτό το σημείο αποτελεί ένα είδος Εξόρυξης Δεδομένων.

Τα περισσότερα πανεπιστήμια επενδύουν μεγάλα χρηματικά ποσά ώστε να καταφέρουν να αποθηκεύσουν τον τεράστιο όγκο δεδομένων που παράγεται. Οι αποθήκες δεδομένων αποτελούν το πιο δημοφιλή χώρο αποθήκευσης εκπαιδευτικών δεδομένων. Αποτελούν χώρο υποδοχής δεδομένων που προέρχονται από ετερογενείς βάσεις δεδομένων. Τα τελευταία χρόνια παρατηρείται η αδυναμία των παραδοσιακών αποθηκών δεδομένων να διαχειριστούν τα Big Data που παράγουν οι φοιτητές, οι ερευνητές και οι καθηγητές των πανεπιστημίων(Santoso, Yulia, 2017). Για την αντιμετώπιση του παραπάνω ζητήματος δημιουργήθηκαν σύγχρονες αποθήκες δεδομένων, ώστε να βελτιώσουν τη διαδικασία λήψης αποφάσεων. Χαρακτηριστικά, έχει σχεδιαστεί ένα Πληροφοριακό Σύστημα που φιλοξενεί την πλατφόρμα Hadoop, η οποία εκτελεί ανάλυση Big Data με μεγάλη ταχύτητα και απόδοση. Λόγω των πολλών ταυτόχρονων χρηστών και του τεράστιου όγκου δεδομένων που διαχειρίζονται, οι σύγχρονες αποθήκες δεδομένων παρέχουν παράλληλη και κατανεμημένη επεξεργασία δεδομένων, με τη βοήθεια του Hadoop.

Τέλος, θα αναφέρουμε την Εκπαιδευτική Αποθήκη. Αποτελεί μία modular πλατφόρμα που παρέχει local και cloud υπηρεσίες, και είναι πρόταση των Amo, Gómez , Hernández-Ibáñez και Fonseca. Παρέχει πολλά επίπεδα ασφάλειας των εκπαιδευτικών δεδομένων που αποθηκεύονται γνωρίζοντας τις ιδιαιτερότητες τους. Σε γενικές γραμμές, λειτουργεί ως αποθήκη δεδομένων με εξουσιοδοτημένη πρόσβαση. Οι χρήστες ανάλογα με το ρόλο τους(καθηγητές, ερευνητές, εκπαιδευτικοί φορείς) έχουν διαφορετικό ρόλο στο σύστημα, δηλαδή άλλα δικαιώματα χρήσης του συγκεκριμένου υβριδικού μοντέλου. Επίσης, εκτός από ασφάλεια και απόρρητο, παρέχει μοντέρνες και αποδοτικές τεχνικές διαχείρισης-επεξεργασίας των δεδομένων. Συνεργάζεται με εφαρμογές τύπου Moodle και με πλατφόρμες όπως το Kahoot, που έχουμε αναφέρει παραπάνω. Τα δεδομένα αποθηκεύονται στη μονάδα αποθήκευσης, αφού πρώτα περάσουν από το στάδιο της προ-επεξεργασίας. Σύμφωνα με τους δημιουργούς της Εκπαιδευτικής Αποθήκης, οι υπηρεσίες που παρέχει μπορούν να βοηθήσουν σε ζητήματα που εμφανίζονται στο χώρο της Εξόρυξης Εκπαιδευτικών Δεδομένων και της Μαθησιακής Αναλυτικής.

***Κεφάλαιο 4:*** ***Ανάλυση προκλήσεων και προοπτικών στον τομέα της μαθησιακής αναλυτικής στην εκπαίδευση.***

Αρχικά, για την εκπόνηση της εργασίας πραγματοποιήθηκε εκτενής ανάγνωση papers και άρθρων από διάφορους συγγραφείς, με στόχο την κατανόηση της έννοιας Μαθησιακή Αναλυτική. Αντιληφθήκαμε πως λειτουργεί στον πραγματικό κόσμο, ποια βήματα απαιτούνται για να θεωρηθεί πλήρης η διαδικασία και ποιες μορφές εκπαιδευτικών δεδομένων χρησιμοποιεί. Όλα τα παραπάνω αποτυπώθηκαν στο Κεφάλαιο 1 της εργασίας μας. Ο συγκεκριμένος αναπτυσσόμενος κλάδος(Μαθησιακή Αναλυτική) αποτελεί αναμφίβολα ένα ισχυρό εργαλείο για τη βελτίωση του Εκπαιδευτικού Συστήματος.. Αναλυτικότερα, σύμφωνα με μελέτες, αναζητά νέους και πιο αποδοτικούς τρόπους με σκοπό την ανάλυση εκπαιδευτικών δεδομένων και την παραγωγή γνώσης, που θα οδηγήσει πιθανώς στην προστασία και την εξέλιξη της διαδικασίας της μάθησης. Αξίζει να τονίσουμε το γεγονός ότι η αξιοποίηση των εκπαιδευτικών δεδομένων μεγάλης κλίμακας που παράγονται και συλλέγονται προσφέρει οφέλη σε σύσσωμη την εκπαιδευτική κοινότητα. Δηλαδή, επηρεάζονται οι μαθητές/φοιτητές, οι καθηγητές, τα σχολεία/πανεπιστήμια, το Υπουργείο Παιδείας και οι συσχετιζόμενοι με το θέμα ερευνητές. Οπότε, καταλαβαίνουμε ότι η Μαθησιακή Αναλυτική και τα αποτελέσματα της έχουν τη δυνατότητα να αλλάξουν αρκετά τη μορφή της μάθησης που γνωρίζουμε μέχρι σήμερα. Ωστόσο, οι ερευνητές και φορείς που σχετίζονται με την εκπαίδευση τονίζουν ότι εκτός από τις προοπτικές που προσφέρει η Μαθησιακή Αναλυτική πρέπει να ληφθούν σοβαρά υπόψη και οι προκλήσεις που εμφανίζονται. Στη συνέχεια του Κεφαλαίου θα γίνει ανάλυση τόσο των προοπτικών της Μαθησιακής Αναλυτικής στην Εκπαίδευση, όσο και αναφορά των ανησυχιών-φόβων της επιστημονικής κοινότητας σχετικά με τον τρόπο που χρησιμοποιείται και τις αρνητικές επιπτώσεις που είναι δυνατόν να παρουσιάσει.

Ξεκινώντας με την ανάλυση των προκλήσεων της ΜΑ στον τομέα της εκπαίδευσης, οι επιστήμονες αναφέρουν τα ζητήματα ασφάλειας δεδομένων. Στην εποχή μας, τόσο στην εκπαίδευση, όσο και σε άλλους τομείς γίνεται καθημερινά συζήτηση μεταξύ απλών ανθρώπων, αλλά και σε επιστημονικούς κύκλους, για τον τρόπο που γίνεται η επεξεργασία των δεδομένων, από ποιους αναλύονται και που αποθηκεύονται. Φόβο δηλαδή των επιστημόνων αποτελεί ο κίνδυνος της διέρευσης των προσωπικών δεδομένων μαθητών και καθηγητών και η χρήση τους από μη-κατάλληλους ανθρώπους για σκοπούς που θα μπορούσαν να κριθούν ως επικίνδυνοι. Όπως αναφέρεται σε σχετική έρευνα, υπάρχουν πολύ σημαντικά νομικά ζητήματα που πρέπει να αντιμετωπισθούν, ώστε να πάρει ένα πανεπιστήμιο άδεια χρήσης δεδομένων φοιτητών(Greller,Drechsler,2012). Είναι λοιπόν λογικό το γεγονός ότι γονείς, μαθητές και καθηγητές ανησυχούν για το απόρρητο τους στο Διαδίκτυο. Για αυτό το λόγο το Υπουργείο Παιδείας των Η.Π.Α το 2012 ανέφερε ότι θα γίνουν σεβαστοί οι κανονισμοί FERPA(Ομοσπονδιακός νόμος για τα οικογενειακά εκπαιδευτικά δικαιώματα), με στόχο την προστασία των δικαιωμάτων των μαθητών. Επίσης, είναι απαραίτητη η διαδικασία ελέγχου των δεδομένων που συλλέγονται, ώστε να μην παραβιάζεται το δικαίωμα της ιδιωτικότητας και να μην επιτρέπεται η συλλογή αποκλειστικά προσωπικών και ευαίσθητων πληροφοριών, ειδικά στην περίπτωση των σχολείων που εμπλέκονται ανήλικοι και πιθανώς μη-εξοικειωμένοι με διαδικτυακές εφαρμογές χρήστες. Αξίζει να αναφέρουμε ότι στη σημερινή εποχή οι σχετικοί νόμοι έχουν γίνει απόλυτα αυστηροί και όλοι ανεξαρτήτως γνώσεων και επιστημονικού υποβάθρου οφείλουν να συμμορφώνονται με εκείνους. Επιπροσθέτως, ίσως είναι αρκετά αγχωτικό για μαθητές και φοιτητές να γνωρίζουν ότι κάθε εκπαιδευτική πράξη που πραγματοποιούν κοινοποιείται άμεσα στους καθηγητές τους. Το συγκεκριμένο ζήτημα ενδιαφέρει άμεσα τη Μαθησιακή Αναλυτική καθώς σχετίζεται με έναν τομέα με τον οποίο συνορεύει και συνεργάζεται, τη Ψυχολογία. Για το Υπουργείο Παιδείας κάθε χώρας και τους επιστήμονες η ψυχολογία των μαθητών και των φοιτητών οφείλει να αποτελεί μείζον ζήτημα(Dietz-Uhler,Hurn,2013). Οπότε, σε αυτό το σημείο γίνεται απαραίτητη η συμμετοχή του εκπαιδευτικού στη διαδικασία, καθώς εκτός από τις επιδόσεις των μαθητών, δηλαδή απομονωμένα νούμερα, πρέπει να λαμβάνονται υπόψη και προσωπικά ή οικογενειακά ζητήματα του κάθε μαθητή ή φοιτητή. Το παραπάνω αντιτίθεται με επιστημονικές απόψεις που προτείνουν τη δημιουργία προφίλ μαθητών και το μετέπειτα χωρισμό τους σε κατηγορίες, ανάλογα με τις επιδόσεις τους σε εργασίες και διαγωνίσματα. Αναντίρρητα, μια τέτοια πολιτική είναι πιθανό να επιφέρει σημαντικότατα προβλήματα στο ψυχισμό των μαθητών και το κοινωνικό τους προφίλ. Δηλαδή, σίγουρα πρέπει να είναι γνωστές οι αποδόσεις τους, ώστε να υπάρχει επίλυση αποριών από μεριάς των καθηγητών σε συγκεκριμένες μαθησιακές ενότητες, ωστόσο πρέπει να αποφευχθούν καταστάσεις περιθωριοποίησης που θυμίζουν παλαιές «σκοτεινές» εποχές. Καταλήγουμε στο ότι πρέπει οι επιστήμονες που θα σχεδιάσουν το εκάστοτε σύστημα ΜΑ να σκεφτούν και εκτός Πληροφορικής, λαμβάνοντας υπόψιν τους τη σχέση μαθητή-καθηγητή και των μαθητών μεταξύ τους. Τέλος, στο πρώτο Κεφάλαιο αναφερθήκαμε σε εξατομικευμένο περιεχόμενο, δηλαδή λύσεις που παρέχονται για κάθε μαθητή/φοιτητή ξεχωριστά έχοντας ως δεδομένα συγκεκριμένες παραμέτρους προσωπικού χαρακτήρα. Από τα παραπάνω καταλαβαίνουμε ότι η εξατομίκευση βρίσκεται κατά κάποιο τρόπο ενάντια στο απόρρητο(Kobsa,2007).

Μια αρκετά μεγάλη πρόκληση που αντιμετωπίζει η Μαθησιακή Αναλυτική στην Εκπαίδευση σχετίζεται με τη μάθηση/διδασκαλία. Έχουμε αναφέρει(στην Εισαγωγή της εργασίας) τη σημασία της μάθησης και γενικότερα της εκπαίδευσης για τον άνθρωπο. Ειδικότερα, είναι σημαντικός ο ρόλος της εκπαίδευσης στη διαδικασία της ανάλυσης των δεδομένων, καθώς είναι πρέπον η εκπαίδευση να καθοδηγεί τη ΜΑ, και όχι το αντίθετο(Greller,Drechsler,2012). Εξηγώντας το παραπάνω, γνωρίζουμε ότι στην εποχή μας υπάρχουν ικανότατοι επιστήμονες και εργαλεία, ώστε να επιτυγχάνεται ανάλυση δεδομένων με τη βοήθεια πολυάριθμων μεθόδων και τεχνικών. Άρα, σε επίπεδο τεχνολογίας έχουν γίνει πάρα πολλά βήματα και συνεχίζουν να δημιουργούνται καθημερινά νέα εργαλεία και να βελτιστοποιούνται τα ήδη υπάρχοντα. Ωστόσο, η ΜΑ οφείλει να συνδυάζει την τεχνολογία με την εκπαίδευση.

Στο σημείο αυτό θα αναφέρουμε μια αρκετά σημαντική πρόκληση που καλείται να αντιμετωπίσει η ΜΑ, ειδικά στη χώρα μας. Αναφερόμαστε στους οικονομικούς πόρους που απαιτούνται από κάθε πανεπιστήμιο ξεχωριστά ή από το Υπουργείο Παιδείας και τη Κυβέρνηση συνολικά, ώστε να ενταχθεί η ΜΑ στα ελληνικά σχολεία και πανεπιστήμια. Όλοι γνωρίζουμε τη σημασία της εκπαίδευσης για τους ανθρώπους στη χώρα μας και πόσο θετικοί θα ήταν όλοι οι συσχετιζόμενοι φορείς σε προσπάθειες βελτίωσης των συνθηκών μάθησης μέσω της ΜΑ. Από την άλλη πλευρά, το μεγάλο οικονομικό κόστος που πρέπει να καλυφθεί αποτελεί τροχοπέδη. Όπως παρατηρήσαμε σε ορισμένα papers ξένων συγγραφέων, στις περισσότερες περιπτώσεις απευθύνονται σε εταιρίες τεχνολογίας και μέσω διαγωνισμού αγοράζουν ένα σύστημα Μαθησιακής Αναλυτικής. Αναντίρρητα, όσο πιο αποδοτικό και δοκιμασμένο είναι το σύστημα ΜΑ τόσο μεγαλύτερο είναι και το κόστος που πρέπει να δαπανηθεί για την αγορά αλλά και τη συντήρησή του. Επίσης, είναι κατανοητό ότι στην κουβέντα «εισέρχεται» ξανά το ζήτημα της προστασίας των δεδομένων. Από τη στιγμή που το εκάστοτε πανεπιστήμιο συνεργάζεται με έναν εξωτερικό φορέα για την παροχή του συστήματος ΜΑ, οι πληροφορίες φοιτητών και καθηγητών κοινοποιούνται και σε εξωτερικούς(σε σχέση με το πανεπιστήμιο) φορείς. Οπότε, όσο πιο αποδοτικό είναι το σύστημα ΜΑ και κάνοντας βήματα προς τα πίσω, τόσο ακριβέστερες είναι οι προβλέψεις που έχει πραγματοποιήσει. Άρα, έχουν έρθει σε επαφή με δεδομένα μεγάλης κλίμακας και εμφανίζεται μεγάλη πιθανότητα ακούσιας αποκάλυψης εκπαιδευτικών δεδομένων. Παράλληλα, εάν θεωρήσουμε ως δεδομένο ότι τα ελληνικά πανεπιστήμια και σχολεία κατάφεραν να βρουν τους πόρους για να αγοράσουν κάποιο σύστημα ΜΑ, τότε προκύπτει μια νέα πρόκληση. Αναφερόμαστε στην κατάρτιση του εκπαιδευτικού προσωπικού, δηλαδή τη δυνατότητα χρήσης του εργαλείου ΜΑ, εκ μέρους των εκπαιδευτικών. Σίγουρα, πρέπει να διοργανωθούν σεμινάρια και να δοθεί οδηγός χρήσης της εφαρμογής σε όλους τους καθηγητές, ώστε μετά από κάποιο χρονικό διάστημα να εφαρμόζουν πρακτικές ΜΑ με ευκολία. Δυστυχώς, μπορεί να εμφανιστούν και πιο σύνθετα προβλήματα, όπως την περίοδο του Covid με τα διαδικτυακά μαθήματα. Δε θεωρείται ακόμη για όλους τους εκπαιδευτικούς απλή η χρήση ενός ηλεκτρονικού υπολογιστή, πόσο μάλλον σύνθετων εφαρμογών ανάλυσης δεδομένων.

Σε πολλά σημεία τα εργασίας μας αναφέραμε τη δυσκολία διαχείρισης δεδομένων μεγάλης κλίμακας(Big Data).Τα εκπαιδευτικά δεδομένα συλλέγονται από ετερογενείς πηγές. Επίσης τα δεδομένα είναι αρκετά πολύπλοκα. Οπότε, είναι επιβεβλημένη η διαχείριση και η ανάλυση των εκπαιδευτικών δεδομένων. Στην περίπτωση των διαδικτυακών μαθημάτων η κατάσταση είναι ακόμα δυσκολότερη(Dringus,2012). Αυτό έχει ως συνέπεια την αδυναμία σε πολλές περιπτώσεις του καθηγητή να αξιολογήσει τον πρόοδο ενός μαθητή/φοιτητή. Το παραπάνω συμβαίνει καθώς τα εκπαιδευτικά δεδομένα που δημιουργούνται μέσω των διαδικτυακών μαθημάτων υστερούν σε διαφάνεια και ενδέχεται να οδηγήσουν σε κακή λήψη αποφάσεων σε σχέση με την πρόοδο των μαθητών/φοιτητών, και η χρήση της ΜΑ να χαρακτηριστεί επιβλαβής. Είναι βέβαιο το γεγονός ότι ορισμένες εξετάσεις μαθημάτων που πραγματοποιούνται εξ αποστάσεως δε θα μπορούσαν με βεβαιότητα να χαρακτηριστούν αδιάβλητες και με αυτό τον τρόπο να οδηγήσουν στη δημιουργία «μπερδεμένων» δεδομένων. Ακόμη, είναι δύσκολο για τον καθηγητή να αντιληφθεί εάν ο κάθε μαθητής/φοιτητής ξεχωριστά και γενικότερα το σύνολο της διαδικτυακής τάξης κατανόησε πλήρως το προς παράδοση θέμα, καθώς σύμφωνα με μελέτες φοιτητές και μαθητές εκφράζουν λιγότερες και πιο τυποποιημένες απορίες σε σύγκριση με τη δια ζώσης διδασκαλία. Συμπερασματικά, από την παραπάνω περιγραφή κατανοούμε ότι η επεξεργασία και η ανάλυση των εκπαιδευτικών δεδομένων που προέρχονται από εξ αποστάσεως διδασκαλία εμφανίζουν μεγαλύτερη πολυπλοκότητα, και αποτελούν σοβαρό ζήτημα.

Συνεχίζουμε με τις προοπτικές που δημιουργούνται από τη χρήση της Μαθησιακής Αναλυτικής στην Εκπαίδευση. Την τελευταία δεκαετία παρατηρείται ότι η ΜΑ εφαρμόζεται στην εκπαίδευση για σκοπούς συνεχούς βελτίωσης του εκπαιδευτικού συστήματος. Δηλαδή, την εποχή της ψηφιακής επανάστασης σημειώθηκε μια στροφή προς την εφαρμογή τεχνολογιών ανάλυσης δεδομένων στον τομέα των ανθρωπιστικών επιστημών(Mustafina,Galiullin,Alloghani,Khanova,2021). Αξίζει να τονιστεί ότι υπάρχουν αρκετά εργαλεία-εφαρμογές ανάλυσης δεδομένων στην αγορά. Ωστόσο, τόσο ορισμένες από τις προκλήσεις που αναφέραμε παραπάνω, όσο και ο γενικότερος χαρακτήρας του εκπαιδευτικού συστήματος έχουν σταθεί εμπόδια στην ανάπτυξη της ΜΑ στην εκπαίδευση. Για την ανάλυση των προοπτικών της ΜΑ στην εκπαίδευση αποφασίσαμε να τις χωρίσουμε ανάλογα με τα υποκείμενα στα οποία αντανακλούν. Αρχικά, όσον αφορά τους μαθητές και τους φοιτητές, μέσω της εφαρμογής της ΜΑ στην εκπαίδευση είναι δυνατή η πρόβλεψη πιθανής πτωτικής τους πορείας και η άμεση παρέμβαση του καθηγητή προς όφελος τους. Επίσης, μέσω εφαρμογής δίνεται η δυνατότητα να καταγράφονται οι απουσίες των μαθητών/φοιτητών και να ειδοποιούνται άμεσα οι γονείς τους. Ακόμη, υπάρχει η δυνατότητα αποστολής αποριών στον καθηγητή μέσω ενός πλαισίου κειμένου(έτσι λειτουργούν οι περισσότερες εφαρμογές ΜΑ στις Η.Π.Α και την Αυστραλία). Γενικότερα, η ΜΑ καλείται να δώσει λύσεις σε ζητήματα που ταλαιπωρούν την εκπαίδευση πολλές δεκαετίες και χωρίς τη χρήση της τεχνολογίας τα συγκεκριμένα προβλήματα πιθανώς θα κινούνταν στην αφάνεια. Επιπροσθέτως, επιτρέπεται ο συνδυασμός διαφορετικών μορφών αξιολόγησης και οι απαντήσεις των μαθητών αναλύονται με ειδικό εργαλείο, αποκαλύπτοντας σημαντικές πληροφορίες που αφορούν την πρόοδό τους. Σύμφωνα με papers που μελετήσαμε, είναι ευκολότερη από πολλές απόψεις, αλλά και συνηθέστερη η χρήση λειτουργιών ΜΑ στα πανεπιστήμια. Ωστόσο, η σημερινή γενιά μαθητών είναι σε μεγάλο βαθμό εξοικειωμένη με διάφορες εφαρμογές, οπότε είναι η κατάλληλη στιγμή για την είσοδο της ΜΑ και στα σχολεία. Ίσως η μεγαλύτερη προοπτική εμφανίζεται στον τρόπο με τον οποίο ο μαθητής/φοιτητής συλλέγει τη γνώση που δέχεται και πως την επεξεργάζεται. Αναλυτικότερα, υπάρχει η δυνατότητα της εξατομικευμένης μάθησης, ώστε να ανταποκριθεί η διαδικασία της διδασκαλίας, και ο καθηγητής φυσικά, στις ατομικές ανάγκες του κάθε μαθητή/φοιτητή(Paolucci,Vancini, Bex II, Cavanaugh,Salama,de Araujo,2024). Με αυτό τον τρόπο παρέχονται ίσες ευκαιρίες μάθησης και δίνεται η δυνατότητα στο μαθητή να ανακαλύψει τα ταλέντα του, και να βελτιώσει τις όποιες αδυναμίες του. Ακόμη, θα ήταν χρήσιμη η οπτικοποίηση συγκεκριμένων δεδομένων με όμορφο τρόπο, ώστε ο μαθητής να κατανοήσει το μαθησιακό του επίπεδο με σκοπό να βελτιωθεί, παρέχοντας όμως όλες τες δικλείδες ασφαλείας ώστε να μην απογοητευτεί. Με βάση τα παραπάνω ενισχύεται η αποτελεσματικότητα της εκπαίδευσης, μεγαλώνει το ενδιαφέρον του μαθητή/φοιτητή για μάθηση και απομακρύνεται ο κίνδυνος της αποχής από τις εκπαιδευτικές δραστηριότητες.

Στη συνέχεια θα επικεντρωθούμε στους καθηγητές και το ρόλο τους στην τάξη σε περίπτωση πλήρους λειτουργίας της Μαθησιακής Αναλυτικής στο εκπαιδευτικό σύστημα. Αρχικά, είναι γνωστό ότι οι καθηγητές μέχρι και σήμερα είναι υπεύθυνοι για μεγάλο όγκο μαθητών. Μέσω της ΜΑ, γίνεται ένας κατά κάποιο τρόπο διαχωρισμός σε πολλά επίπεδα. Δηλαδή, επιτρέπεται στον καθηγητή να κάνει ενισχυτικά μαθήματα σε κάποιους μαθητές για συγκεκριμένο κεφάλαιο της ύλης, που εμφανίζουν κενά. Επίσης, κάτι ανάλογο μπορεί να πραγματοποιηθεί όσον αφορά τις εξετάσεις. Οπότε, με αυτό τον τρόπο και ενεργώντας πολύ προσεκτικά ο διδάσκων παρέχει εξατομικευμένη μάθηση. Παράλληλα, είναι δυνατή η μείωση του φόρτου εργασίας του καθηγητή, με τη βοήθεια της τεχνολογίας. Σίγουρα η «παπαγαλία» δε βοηθάει τη νοητική ανάπτυξη των μαθητών και τη γενικότερη διαδικασία της μάθησης. Παρόμοια αποτελέσματα έχει και η διόρθωση πολυάριθμων διαγωνισμάτων που αποτελούνται από πολλές σελίδες, εκ μέρους των καθηγητών. Το θέμα είναι, και αναμφισβήτητα όλοι οι εμπλεκόμενοι με την εκπαίδευση φορείς συμφωνούν, να βελτιωθούν οι συνθήκες μάθησης και να αυξηθούν τα επίπεδα ικανοποίησης μαθητών και καθηγητών τις ώρες που ασχολούνται με θέματα εκπαίδευσης. Επιπροσθέτως, η είσοδος της ΜΑ στην Εκπαίδευση θα βοηθήσει στην αυτοκριτική των καθηγητών. Μέσω της εφαρμογής παρέχονται πληροφορίες σχετικά με το πλάνο του εκάστοτε καθηγητή και συγκεκριμένες παρατηρήσεις, με σκοπό τη βελτίωση των μαθησιακών πρακτικών του. Εκτός από τους μαθητές/φοιτητές είναι απαραίτητο να αξιολογούνται και οι καθηγητές.

Ακολουθεί ένα πολύ σημαντικό ζήτημα. Είναι δυστυχώς σύνηθες φαινόμενο ορισμένοι καθηγητές να ασχολούνται περισσότερο με ορισμένους μαθητές, και κυρίως να αποθαρρύνουν με τον τρόπο τους, τους μαθητές που αντιμετωπίζουν κενά στα μαθήματα. Όμως, με τη χρήση της ΜΑ οι παραπάνω ανισορροπίες και τακτικές περιθωριοποίησης θα αντιμετωπισθούν. Αξίζει να τονιστεί ότι σε αντικειμενικά δύσκολα μαθήματα στο επίπεδο του σχολείου, όπως τα Μαθηματικά, γίνεται αξιολόγηση των γνώσεων των μαθητών από προηγούμενα έτη. Έτσι, μέσω του εργαλείου e-learning που χρησιμοποιεί το σχολείο, οι μαθητές που το χρειάζονται έρχονται σε επαφή με πρόσθετο υλικό και λύνουν αυτοματοποιημένα quiz, δίνοντας τους την ευκαιρία να αποτυπώσουν στο μυαλό τους γνώσεις που έχουν διδαχθεί στο παρελθόν(Paolucci,Vancini,Bex II,Cavanaugh,Salama,de Araujo,2024). Εν κατακλείδι, κρίνεται απαραίτητη η αποστολή σχολίων και προτεινόμενων βελτιώσεων που αφορούν την εφαρμογή e-learning που χρησιμοποιεί το εκάστοτε σχολείο/πανεπιστήμιο από τους καθηγητές προς τους υπεύθυνους της εφαρμογής. Αναμφίβολα, ο καθηγητής έχει καθαρή εικόνα για τους μαθητές, τη διδασκαλία και τη διαδικασία της ΜΑ. Επίσης, λαμβάνει το feedback από τη χρήση της εφαρμογής μέσω των μαθητών και μπορεί να συγκρίνει την πρόοδό τους σε συγκεκριμένα θέματα με την απόδοσή τους την εποχή πριν τη χρήση της ΜΑ.

Όπως έγινε κατανοητό και από την ανάλυση των προκλήσεων στον τομέα της ΜΑ στην Εκπαίδευση, φαίνεται ότι υπάρχουν αντικρουόμενες έρευνες σχετικά με το συγκεκριμένο θέμα. Εκτός από την ανάπτυξη νέων ερευνητικών μεθόδων στον τομέα της εκπαίδευσης πρέπει να δημιουργηθεί στοχευμένη έρευνα για την απόδοση της ΜΑ στην εκπαίδευση. Αναλυτικότερα, οι ερευνητές οφείλουν να βοηθήσουν το Υπουργείο Παιδείας. εκπαιδευτικούς συμβούλους, καθηγητές και μαθητές/φοιτητές να κατανοήσουν τη σημασία της ΜΑ στην εκπαίδευση. Επίσης, ως προοπτική θα μπορούσε να χαρακτηριστεί η συλλογή χρήσιμων δεδομένων που αφορούν σχολεία και πανεπιστήμια που ήδη χρησιμοποιούν το συγκεκριμένο σύστημα. Το παραπάνω είναι χρήσιμο και για τους επιστήμονες που δημιουργούν τα εργαλεία/εφαρμογές ΜΑ, αλλά και για αυτούς που τα συντηρούν. Αναμφίβολα, για να αντιμετωπισθούν οι προκλήσεις και οι κίνδυνοι που «αναδύονται» είναι πολύ σημαντικός ο προσωπικός σχεδιασμός και η υλοποίηση της εφαρμογής ΜΑ που θα χρησιμοποιούν καθημερινά μαθητές και καθηγητές. Όσο καλύτερο είναι το εργαλείο που χρησιμοποιούν, τόσο θετικότερος θα είναι και ο αντίκτυπος του στη μάθηση. Αναμφισβήτητα, για το παραπάνω είναι επιβεβλημένη η συνεργασία των ερευνητών με τους εκπαιδευτικούς φορείς. Ο καθένας γνωρίζει καλύτερα τον τομέα του. Ωστόσο, χωρίς τη συνεργασία τους είναι αδύνατη η αντιμετώπιση των προκλήσεων και της καχυποψίας που ίσως προκύψει. Τέλος η ευθύνη του Υπουργείου Παιδείας είναι αρκετά μεγάλη. Για να επιτευχθεί αποτελεσματική ανάλυση της πορείας μαθητών και φοιτητών, των πρακτικών των καθηγητών και της απόδοσης της εφαρμογής e-learning πρέπει οι εκπαιδευτικοί φορείς να συντονίζουν το συνολικό σύστημα, να λαμβάνουν αποφάσεις ελαχιστοποιώντας το εκτιμώμενο ρίσκο, και να έχουν στο μυαλό τους ότι η μάθηση είναι το κέντρο όλου του συστήματος. Με τη συνεργασία όλων των παραπάνω, είναι πιθανή η βελτίωση της Μαθησιακής Αναλυτικής στην Εκπαίδευση και η καθιέρωσή της ως απαραίτητο εργαλείο για το Εκπαιδευτικό Σύστημα.

Βιβλιογραφικές Αναφορές

**1.** Cristobal Romero, Sebastian Ventura(2013). “Educational Data mining and Learning Analytics: An updated survey”. [2402.07956.pdf (arxiv.org)](https://arxiv.org/ftp/arxiv/papers/2402/2402.07956.pdf)

**2.** Santosh Kumar Ray, Mohammed Saeed(2018). “Applications of Educational Data Mining and Learning Analytics Tools in Handling Big Data in Higher Education: Trends, Issues, and Challenges”. [(PDF) Applications of Educational Data Mining and Learning Analytics Tools in Handling Big Data in Higher Education: Trends, Issues, and Challenges (researchgate.net)](https://www.researchgate.net/publication/326554623_Applications_of_Educational_Data_Mining_and_Learning_Analytics_Tools_in_Handling_Big_Data_in_Higher_Education_Trends_Issues_and_Challenges)

**3.** Marie Bienkowski, Mingyu Feng, Barbara Means(2012). “Enhancing Teaching and Learning Through Educational Data Mining and Learning Analytics: An Issue Brief”. [ED611199.pdf](https://files.eric.ed.gov/fulltext/ED611199.pdf)

**4.** Ryan S. Baker, Paul Salvador Inventado(2016).”Educational Data Mining and Learning Analytics”. [Educational Data Mining and Learning Analytics (edtechbooks.org)](https://edtechbooks.org/lidtfoundations/educational_data_mining_and_learning_analytics)

**5.** Beth Dietz-Uhler, Janet E. Hurn(2013). “Using Learning Analytics to Predict (and Improve) Student Success: A Faculty Perspective”. [12.1.2.pdf (ncolr.org)](https://www.ncolr.org/jiol/issues/pdf/12.1.2.pdf)

**6.** Jamila Mustafina, Mohamed Ahmed Alloghani, Lenar Galiullin, Aygul Khanova (2021). “LEARNING ANALYTICS: CHALLENGES AND PERSPECTIVES”. [(PDF) LEARNING ANALYTICS: CHALLENGES AND PERSPECTIVES (researchgate.net)](https://www.researchgate.net/publication/350417862_LEARNING_ANALYTICS_CHALLENGES_AND_PERSPECTIVES)

**7.** Catherine Paolucci, Sam Vancini, Richard T. Bex II, Catherine Cavanaugh, Christine Salama, Zandra de Araujo(2024). “A review of learning analytics opportunities and challenges for K-12 education”. [A review of learning analytics opportunities and challenges for K-12 education - ScienceDirect](https://www.sciencedirect.com/science/article/pii/S2405844024017985)

**8.** Yupei Zhang, Yue Yun, Rui An, Jiaqi Cui, Huan Dai , Xuequn Shang(2021). “Educational Data Mining Techniques for Student Performance Prediction: Method Review and Comparison Analysis”. [Educational Data Mining Techniques for Student Performance Prediction: Method Review and Comparison Analysis - PMC (nih.gov)](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8688359/)

**9.** Ryan S.J.d. Baker(2010). “Data Mining for Education”. [https://learninganalytics.upenn.edu/ryanbaker/Encyclopedia%20Chapter%20Draft%20v10%20-fw.pdf](https://learninganalytics.upenn.edu/ryanbaker/Encyclopedia%20Chapter%20Draft%20v10%20-fw.pdf#:~:text=There%20are%20a%20wide%20variety%20of%20current%20methods,models%2C%20and%20distillation%20of%20data%20for%20human%20judgment.)

**10.** Myla Arcinas, Guna Sekhar Sajja, Shazia Asif, Sanjeev Gour, Ethelbert Okoronkwo, Mohd Naved(2021). “ROLE OF DATA MINING IN EDUCATION FOR IMPROVING STUDENTS PERFORMANCE FOR SOCIAL CHANGE”. [(PDF) ROLE OF DATA MINING IN EDUCATION FOR IMPROVING STUDENTS PERFORMANCE FOR SOCIAL CHANGE (researchgate.net)](https://www.researchgate.net/publication/352736741_ROLE_OF_DATA_MINING_IN_EDUCATION_FOR_IMPROVING_STUDENTS_PERFORMANCE_FOR_SOCIAL_CHANGE)

**11.** Monika Goyal, Rajan Vohra(2012). “Applications of Data Mining in Higher Education”.

[citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=0fc705ef34776db8f50d64b4b82bbd11ca5ebee5](https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=0fc705ef34776db8f50d64b4b82bbd11ca5ebee5)

**12**. Fatima Alshareef, Hosam Alhakami, Tahani Alsubait, Abdullah Baz(2020). “Educational Data Mining Applications and Techniques”. [Paper\_94-Educational\_Data\_Mining\_Applications.pdf (thesai.org)](https://thesai.org/Downloads/Volume11No4/Paper_94-Educational_Data_Mining_Applications.pdf)

**13.** Jaya Srivastava, Abhay Kumar Srivastava(2021). “Data Mining in Education Sector: A Review”. [Data-Mining-in-Education-Sector-A-Review.pdf (researchgate.net)](https://www.researchgate.net/profile/Abhay-Srivastava/publication/356065560_Data_Mining_in_Education_Sector_A_Review/links/618a9e763068c54fa5c2e322/Data-Mining-in-Education-Sector-A-Review.pdf)

**14.** Elaf Abu Ambrieh, Thair Hamtini, Ibrahim Alijarah(2016). “Mining Educational Data to Predict Student’s academic Performance using Ensemble Methods”. [13.pdf (evo-ml.com)](https://evo-ml.com/ibrahim/publications/13.pdf)

**15.** Cristobal Romero, Sebastian Ventura(2020). “Educational Data Mining and Learning Analytics: An Updated Survey”. [Educational Data Mining and Learning Analytics: An Updated Survey | Request PDF (researchgate.net)](https://www.researchgate.net/publication/338052900_Educational_Data_Mining_and_Learning_Analytics_An_Updated_Survey)

**16.** H. Michael Chung, Paul Gray(1999).Special Section: Data Mining”. [Special Section: Data Mining: Journal of Management Information Systems: Vol 16, No 1 (tandfonline.com)](https://www.tandfonline.com/doi/abs/10.1080/07421222.1999.11518231)

**17.**  Matthew Peacock(1998). “Exploring the gap between teachers' and learners' beliefs about ‘useful’ activities for EFL”. [Exploring the gap between teachers' and learners' beliefs about ‘useful’ activities for EFL - Peacock - 1998 - International Journal of Applied Linguistics - Wiley Online Library](https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1473-4192.1998.tb00131.x)

**18.** Usama M. Fayyad(1996).” Data Mining and Knowledge Discovery in Databases:

Applications in Astronomy and Planetary Science”. [AAAI96-289.pdf](https://cdn.aaai.org/AAAI/1996/AAAI96-289.pdf)

**19.** Bui Duc Trung; Ngo Tung Son; Nguyen Duy Tung; Kieu Anh Son; Bui Ngoc Anh; Phan Truong Lam(2023). “Educational Data Mining: A Systematic Review on the Applications of Classical Methods and Deep Learning Until 2022”. [Educational Data Mining: A Systematic Review on the Applications of Classical Methods and Deep Learning Until 2022 | IEEE Conference Publication | IEEE Xplore](https://ieeexplore.ieee.org/abstract/document/10212273/authors#authors)

**20.** Cristobal Romero, Sebastian Ventura(2007). “Educational data mining: A survey from 1995 to 2005”. [(PDF) Educational data mining: A survey from 1995 to 2005 (researchgate.net)](https://www.researchgate.net/publication/222815211_Educational_data_mining_A_survey_from_1995_to_2005)

**21.** Lap Kei Lee, Simon K.S. Cheung, Lam For Kwok(2020). “Learning analytics: current trends and innovative practices”. [Learning analytics: current trends and innovative practices — Hong Kong Metropolitan University (hkmu.edu.hk)](https://scholars.hkmu.edu.hk/en/publications/learning-analytics-current-trends-and-innovative-practices)

**22.** John P Campbell, Diana G. Oblinger(2012). “Academic Analytics”. [Academic Analytics | Request PDF (researchgate.net)](https://www.researchgate.net/publication/242453132_Academic_Analytics)

**23.** Tan, Steinbach, Karpatne, Kumar. “Εισαγωγή στην Εξόρυξη Δεδομένων”(Βιβλίο από τον Εύδοξο).

**24.** Leskovec, Rajaraman, Ullman. “Εξόρυξη από Μεγάλα Σύνολα Δεδομένων””(Βιβλίο από τον Εύδοξο)..

**25.** Ryan S.J.d. Baker(2008).“ Data Mining for Education”. [Encyclopedia Chapter Draft v10 -fw.pdf (upenn.edu)](https://learninganalytics.upenn.edu/ryanbaker/Encyclopedia%20Chapter%20Draft%20v10%20-fw.pdf)

**26.** G.W. Dekker, M. Pechenizkiy, J.M. Vleeshouwers(2009).“ Predicting students drop out : a case study”. [Predicting students drop out : a case study — Eindhoven University of Technology research portal (tue.nl)](https://research.tue.nl/en/publications/predicting-students-drop-out-a-case-study)

**27.** Rupa Chiramdasu, Gautam Srivastava, Sweta Bhattacharya, Praveen Kumar Reddy, Thippa Reddy Gadekallu(2021).“ Malicious URL Detection using Logistic Regression”. [Malicious URL Detection using Logistic Regression | IEEE Conference Publication | IEEE Xplore](https://ieeexplore.ieee.org/abstract/document/9524269)

**28.** Monika Goyal, Rajan Vohra(2012).“ Applications of Data Mining in Higher Education”. [(PDF) Applications of Data Mining in Higher Education (researchgate.net)](https://www.researchgate.net/publication/264888758_Applications_of_Data_Mining_in_Higher_Education)

**29.** Wolfgang Greeller, Hendrick Drachsler(2012**)**.“ Translating learning into numbers: A generic framework for learning analytics”. [Translating learning into numbers: A generic framework for learning analytics. (apa.org)](https://psycnet.apa.org/record/2012-33665-003)

**30.** Jen-Her Wu, Tzyh-Lih Hsia, Yi-Wen Liao, Robert D. Tennyson(2008).“ What Determinates Student Learning Satisfaction in a Blended E-Learning System Environment?”. [What Determinates Student Learning Satisfaction in a Blended E-Learning System Environment? | Request PDF (researchgate.net)](https://www.researchgate.net/publication/221228948_What_Determinates_Student_Learning_Satisfaction_in_a_Blended_E-Learning_System_Environment)

**31.** William W. Guo(2010).“ Incorporating statistical and neural network approaches for student course satisfaction analysis and prediction”. [Incorporating statistical and neural network approaches for student course satisfaction analysis and prediction - ScienceDirect](https://www.sciencedirect.com/science/article/abs/pii/S095741740900880X)

**32.** Alfred Kobsa(2007).“ Privacy-Enhanced Web Personalization”. [2007-AWBS-privacy-kobsa.pdf (uci.edu)](https://ics.uci.edu/~kobsa/papers/2007-AWBS-privacy-kobsa.pdf)

**33.** L. Drignus(2012).“ Learning Analytics Considered Harmful”. [[PDF] Learning Analytics Considered Harmful. | Semantic Scholar](https://www.semanticscholar.org/paper/Learning-Analytics-Considered-Harmful.-Dringus/e31942c1cbcc629f8f21ec870d81f8af054c74e6)

**34**. Emily Gustafsson-Wright, Sarah Osborne, Muskan Aggarwal(2022). “DIGITAL TOOLS FOR REAL-TIME DATA COLLECTION IN EDUCATION”. [Digital tools for real-time data collection in education | Brookings](https://www.brookings.edu/articles/digital-tools-for-real-time-data-collection-in-education/)

**35.** J. Spencer Clark, Suzanne Porath, Julie Thiele, Morgan Jobe(2020) “Collecting Data in Your Classroom”. [Collecting Data in Your Classroom – Action Research (pressbooks.pub)](https://kstatelibraries.pressbooks.pub/gradactionresearch/chapter/chapt5/)

**36.** Daniel Amo, Pablo Gómez, Luis Hernández-Ibáñez, David Fonseca

(2021). “Educational Warehouse: Modular, Private and Secure Cloudable Architecture System for Educational Data Storage, Analysis and Access”. [Applied Sciences | Free Full-Text | Educational Warehouse: Modular, Private and Secure Cloudable Architecture System for Educational Data Storage, Analysis and Access (mdpi.com)](https://www.mdpi.com/2076-3417/11/2/806)

**37.** Leo Willyanto Santoso, Yulia(2017). “Data Warehouse with Big Data Technology for Higher Education”. [Data Warehouse with Big Data Technology for Higher Education - ScienceDirect](https://www.sciencedirect.com/science/article/pii/S1877050917329022)

**38.** Iris Garner. “Data in Education”. [Data in Education | Learning A-Z](https://www.learninga-z.com/site/resources/breakroom-blog/data-in-education)

**39.** Cristóbal Romero, José Raúl Romero, Sebastian Ventura(2014). “A Survey on Pre-Processing Educational Data”. [(PDF) A Survey on Pre-Processing Educational Data (researchgate.net)](https://www.researchgate.net/publication/277195076_A_Survey_on_Pre-Processing_Educational_Data)

***Πρακτικό Μέρος***

Για το Πρακτικό Μέρος της εργασίας επιλέξαμε να υλοποιήσουμε ένα σενάριο Εξόρυξης Εκπαιδευτικών Δεδομένων και Μαθησιακής Αναλυτικής χρησιμοποιώντας το εργαλείο Tableau. Πρόκειται για μια εφαρμογή οπτικοποίησης και ανάλυσης δεδομένων, δίχως τη συγγραφή κώδικα. Αρχικά πρέπει να βρούμε ένα dataset και να το εισάγουμε στο Tableau. Έπειτα, ανάλογα με το τι θέλουμε να πραγματοποιήσουμε, επιλέγουμε τα κατάλληλα χαρακτηριστικά του συνόλου δεδομένων, προσθέτουμε τα κατάλληλα φίλτρα και δημιουργούμε διαγράμματα. Δίνει τη δυνατότητα δημιουργίας εκατοντάδων διαγραμμάτων, που καλύπτουν τις περισσότερες ανάγκες του χρήστη όσον αφορά την οπτικοποίηση. Επιλέξαμε να χρησιμοποιήσουμε το συγκεκριμένο εργαλείο, καθώς από τα προτεινόμενα εργαλεία της εκφώνησης φαίνεται πως είναι αρκετά εύκολο στη χρήση του και προσφέρει φιλικό user interface.

Αρχικά, αναζητήσαμε datasets που να ανταποκρίνονται με το θέμα της εργασίας μας στον ιστότοπο Kaggle. Καταλήξαμε στο *dataset: Higher Education Students Performance Evaluation*(<https://www.kaggle.com/datasets/csafrit2/higher-education-students-performance-evaluation/data>*)*. Το συγκεκριμένο dataset το έχουν υλοποιήσει οι Nevriye Yilmaz και Boran Sekeroglu, και είναι σε μορφή csv αρχείου(student\_prediction.csv). Οι παραπάνω επιστήμονες συλλέξαν τα δεδομένα το 2019 από τους φοιτητές της Πολυτεχνικής Σχολής και της Σχολής Επιστημών της Αγωγής. Άρα, τα δεδομένα και γενικότερα το σενάριο που θα υλοποιήσουμε στο Tableau αναφέρονται στην Τριτοβάθμια Εκπαίδευση. Το συγκεκριμένο σύνολο δεδομένων, αποτελεί ένα μέρος των σεναρίων και των μεθόδων που περιγράψαμε στο Θεωρητικό Μέρος της εργασίας.

Κατά τη διάρκεια των πειραμάτων ανακαλύψαμε ότι χρειάζεται να κάνουμε αλλαγές στον τύπο δεδομένων ορισμένων χαρακτηριστικών του dataset. Για αυτό και δημιουργήσαμε τα αρχεία updated\_student\_prediction.csv και updated\_student\_prediction2.csv, με τη σημείωση ότι για τη δημιουργία τους χρησιμοποιήσαμε κώδικα σε Python(χρήση της δημοφιλούς βιβλιοθήκης pandas) και ειδικότερα τα αρχεία changes.py και changes2.py αντίστοιχα, τα οποία παραθέτουμε παρακάτω.

Αρχείο changes.py(Μετατροπή των χαρακτηριστικών Notes,Attend\_dept,grade και Impact από αριθμητικά σε κατηγορικά, για να μπορέσουμε να δημιουργήσουμε συγκεκριμένα διαγράμματα που θα παρουσιάσουμε στη συνέχεια):

**import** pandas **as** pd  
  
df = pd.read\_csv(**'student\_prediction.csv'**)  
  
notes\_mapping = {1: **'Never'**, 2: **'Sometimes'**, 3: **'Always'**}  
df[**'NOTES'**] = df[**'NOTES'**].map(notes\_mapping)  
  
attend\_mapping = {1: **'Always'**, 2: **'Sometimes'**, 3: **'Never'**}  
df[**'ATTEND\_DEPT'**] = df[**'ATTEND\_DEPT'**].map(attend\_mapping)  
  
grades\_mapping = {0: **'Fail'**, 1: **'DD'**, 2: **'DC'**,3: **'CC'**, 4: **'CB'**, 5: **'BB'**,6: **'BA'**,7: **'AA'**}  
df[**'GRADE'**] = df[**'GRADE'**].map(grades\_mapping)  
  
projects\_mapping = {1: **'Positive'**, 2: **'Negative'**, 3: **'Neutral'**}  
df[**'IMPACT'**] = df[**'IMPACT'**].map(projects\_mapping)  
  
df.to\_csv(**'updated\_student\_prediction.csv'**, index=**False**)

Αρχείο changes2.py – Προσθήκη κώδικα στο αρχείο changes.py, ώστε να επέμβουμε στο πρωτότυπο αρχείο(Μετατροπή των χαρακτηριστικών cuml\_gpa και exp\_gpa από αριθμητικά σε κατηγορικά):

**import** pandas **as** pd  
  
df = pd.read\_csv(**'student\_prediction.csv'**)  
  
notes\_mapping = {1: **'Never'**, 2: **'Sometimes'**, 3: **'Always'**}  
df[**'NOTES'**] = df[**'NOTES'**].map(notes\_mapping)  
  
attend\_mapping = {1: **'Always'**, 2: **'Sometimes'**, 3: **'Never'**}  
df[**'ATTEND\_DEPT'**] = df[**'ATTEND\_DEPT'**].map(attend\_mapping)  
  
grades\_mapping = {0: **'Fail'**, 1: **'DD'**, 2: **'DC'**,3: **'CC'**, 4: **'CB'**, 5: **'BB'**,6: **'BA'**,7: **'AA'**}  
df[**'GRADE'**] = df[**'GRADE'**].map(grades\_mapping)  
  
projects\_mapping = {1: **'Positive'**, 2: **'Negative'**, 3: **'Neutral'**}  
df[**'IMPACT'**] = df[**'IMPACT'**].map(projects\_mapping)  
  
cgpa\_mapping = {1: **'<2.00'**, 2: **'2.00-2.49'**, 3: **'2.50-2.99'**, 4: **'3.00-3.49'**, 5: **'above 3.49'**}  
df[**'CUML\_GPA'**] = df[**'CUML\_GPA'**].map(cgpa\_mapping)  
  
exgpa\_mapping = {1: **'<2.00'**, 2: **'2.00-2.49'**, 3: **'2.50-2.99'**, 4: **'3.00-3.49'**, 5: **'above 3.49'**}  
df[**'EXP\_GPA'**] = df[**'EXP\_GPA'**].map(exgpa\_mapping)  
  
df.to\_csv(**'updated\_student\_prediction2.csv'**, index=**False**)

Περιγραφή των χαρακτηριστικών του dataset, δηλαδή του πρωτότυπου csv αρχείου. Σε παρένθεση βρίσκονται οι διάφορες τιμές που επιτρέπεται να έχει το κάθε χαρακτηριστικό και με έντονα γράμματα εκείνα τα οποία μας φάνηκαν ενδιαφέροντα και τα εντάξαμε στο σενάριο μας:

0- **Student ID** (unique for each student)

1- Student Age (1: 18-21, 2: 22-25, 3: above 26)

2- Sex (1: female, 2: male)

3- Graduated high-school type: (1: private, 2: state, 3: other)

4- Scholarship type: (1: None, 2: 25%, 3: 50%, 4: 75%, 5: Full)

5- Additional work: (1: Yes, 2: No)

6- Regular artistic or sports activity: (1: Yes, 2: No)

7- Do you have a partner: (1: Yes, 2: No)

8- Total salary if available (1: USD 135-200, 2: USD 201-270, 3: USD 271-340, 4: USD 341-410, 5: above 410)

9- Transportation to the university: (1: Bus, 2: Private car/taxi, 3: bicycle, 4: Other)

10- Accommodation type in Cyprus: (1: rental, 2: dormitory, 3: with family, 4: Other)

11- **Mother's education**: (1: primary school, 2: secondary school, 3: high school, 4: university, 5: MSc., 6: Ph.D.)

12- **Father's education**: (1: primary school, 2: secondary school, 3: high school, 4: university, 5: MSc., 6: Ph.D.)

13- Number of sisters/brothers (if available): (1: 1, 2:, 2, 3: 3, 4: 4, 5: 5 or above)

14- Parental status: (1: married, 2: divorced, 3: died - one of them or both) \*\*\*Listed as "Kids"…woops

15- Mother's occupation: (1: retired, 2: housewife, 3: government officer, 4: private sector employee, 5: self-employment, 6: other)

16- Father's occupation: (1: retired, 2: government officer, 3: private sector employee, 4: self-employment, 5: other)

17- **Weekly study hours**: (1: None, 2: <5 hours, 3: 6-10 hours, 4: 11-20 hours, 5: more than 20 hours)

18- **Reading frequency** (non-scientific books/journals): (1: None, 2: Sometimes, 3: Often)

19- Reading frequency (scientific books/journals): (1: None, 2: Sometimes, 3: Often)

20- Attendance to the seminars/conferences related to the department: (1: Yes, 2: No)

21- **Impact of your projects/activities on your success**: (1: positive, 2: negative, 3: neutral)

22- Attendance to classes (1: always, 2: sometimes, 3: never)

23- Preparation to midterm exams 1: (1: alone, 2: with friends, 3: not applicable)

24- Preparation to midterm exams 2: (1: closest date to the exam, 2: regularly during the semester, 3: never)

25- **Taking notes in classes**: (1: never, 2: sometimes, 3: always)

26- Listening in classes: (1: never, 2: sometimes, 3: always)

27- Discussion improves my interest and success in the course: (1: never, 2: sometimes, 3: always)

28- Flip-classroom: (1: not useful, 2: useful, 3: not applicable)

29- **Cumulative grade point average in the last semester** (/4.00): (1: <2.00, 2: 2.00-2.49, 3: 2.50-2.99, 4: 3.00-3.49, 5: above 3.49)

30- **Expected Cumulative grade point average in the graduation** (/4.00): (1: <2.00, 2: 2.00-2.49, 3: 2.50-2.99, 4: 3.00-3.49, 5: above 3.49)

31- Course ID

32- *OUTPUT:* **Grade** (0: Fail, 1: DD, 2: DC, 3: CC, 4: CB, 5: BB, 6: BA, 7: AA)

**Στη συνέχεια ακολουθεί η παρουσίαση του σεναρίου με σχετικά screenshots και τις κατάλληλες επεξηγήσεις:**

Παράδειγμα 1

Αρχικά θέλουμε να οπτικοποιήσουμε κάτι πολύ απλό, δηλαδή πόσες ώρες διαβάζει ανά εβδομάδα ο κάθε ένας από τους 145 μαθητές που βρίσκονται στο σύνολο δεδομένων που φορτώσαμε στο εργαλείο Tableau.

Οπότε, χρησιμοποιούμε τα χαρακτηριστικά **Student ID** και **Weekly study hours**, δηλαδή τα **STUDENTID** και **STUDY\_HRS**, όπως αναφέρονται στο csv αρχείο. Αξίζει να σημειώσουμε ότι το STUDY\_HRS παίρνει τιμές από 1-5, όμως δεν είναι ώρες αλλά: ***1: None, 2: <5 hours, 3: 6-10 hours, 4: 11-20 hours, 5: more than 20 hours***.

Αναμφίβολα, η γνώση των ωρών που μελετά ένας φοιτητής κάθε εβδομάδα θα βοηθούσε σε μεγάλο βαθμό τον καθηγητή να καταλάβει αρκετά πράγματα. Αναλυτικότερα, πρώτα από όλα, πρέπει να δει ποιο ποσοστό των φοιτητών διαβάζει και στη συνέχεια να το συνδέσει με άλλα χαρακτηριστικά, όπως τους παράγοντες που δεν του επιτρέπουν να διαβάζει(εργάζεται, έχει παιδιά άρα και αρκετά μειωμένο χρόνο) ή γιατί ενώ διαβάζει ικανοποιητικό αριθμό ωρών δεν αποδίδει καλά σε ορισμένες εργασίες ή διαγωνίσματα. Πρέπει σε αυτό το σημείο να προσθέσουμε πως όλα όσα αναφέραμε σίγουρα θα ήθελε να τα γνωρίζει και το εκάστοτε Πανεπιστήμιο, αλλά και το Υπουργείο Παιδείας, ώστε και εκείνα να προβούν στη δική τους ανάλυση και να βελτιώσουν την ποιότητα λήψης αποφάσεων όσον αφορά τα Πανεπιστήμια. Εν κατακλείδι, αναντίρρητα μέσω της ανάλυσης των δεδομένων που παράγει ο κάθε φοιτητής, παρουσιάζονται και ενδιαφέροντα στοιχεία που πρέπει να διαμοιραστούν στους φοιτητές, ώστε να συνειδητοποιήσουν κρυφή γνώση που είναι αρκετά σημαντική για τους ίδιους και για τη φοίτηση τους.

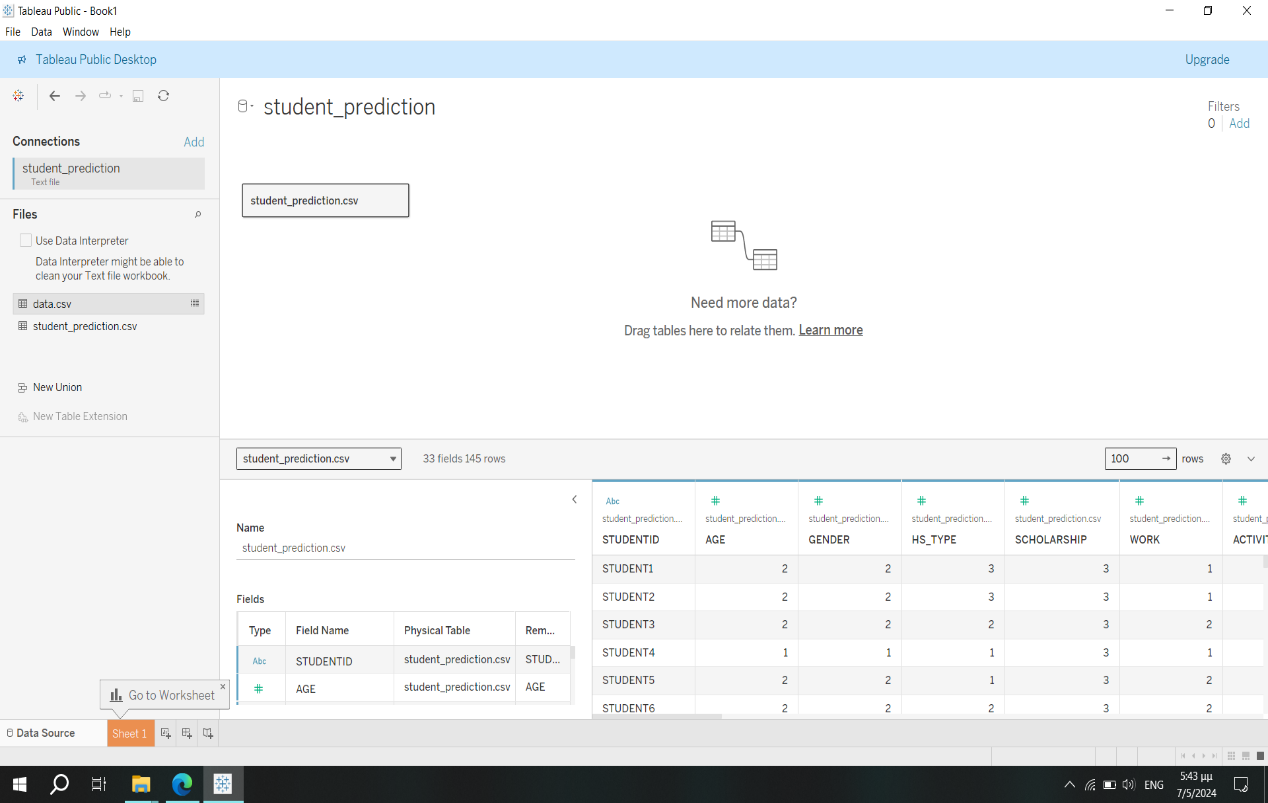
Μπορούμε το διάγραμμα των 2 στηλών(STUDENTID και STUDY\_HRS) να το δημιουργήσουμε πολλούς τρόπους. Επιλέχθηκε η κατηγορία Circle(αντί των Bars, Pie, και πολλών άλλων), η οποία αποτυπώνει το κάθε στιγμιότυπο ως κυκλάκι.

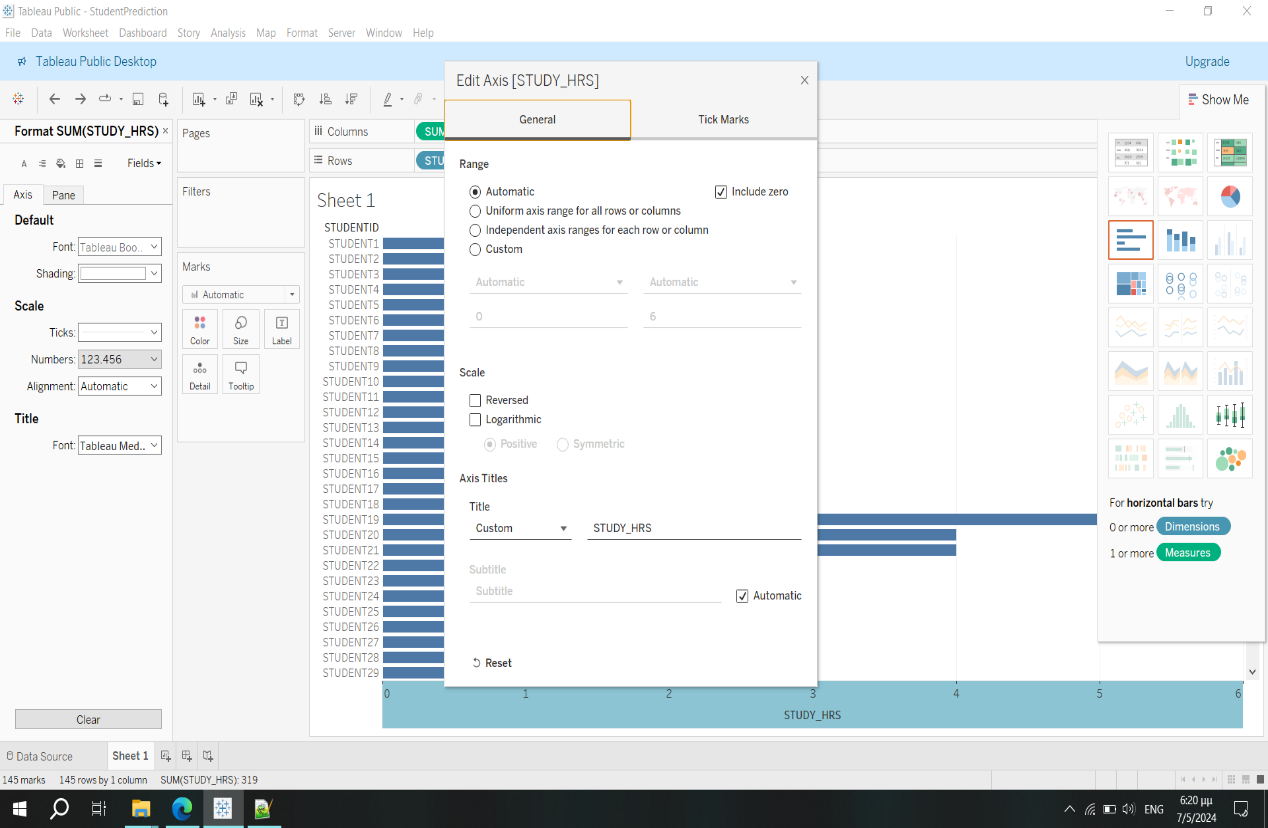
Γενικά τα χαρακτηριστικά με πράσινο χρώμα αποτελούν τα μέτρα, ενώ με μπλε τις διαστάσεις. Επιπροσθέτως, με # συμβολίζονται τα χαρακτηριστικά που έχουν αριθμούς ως τιμές τους, ενώ με Abc αυτά που περιέχουν κατηγορικά χαρακτηριστικά, δηλαδή συμβολοσειρές. Επιπλέον, το πεδίο Filters χρησιμοποιείται για να επιλέξουμε συγκεκριμένες κατηγορίες δεδομένων που θέλουμε οπτικοποιήσουμε, άρα να εστιάσουμε σε συγκεκριμένο υπο-χώρο(που θεωρούμε ότι εμφανίζει αυξημένο ενδιαφέρον) την ανάλυση μας. Φίλτρα προσθέτουμε για συγκεκριμένα χαρακτηριστικά, ανάλογα το σενάριο που θέλουμε να υλοποιήσουμε. Τέλος, το πεδίο Marks είναι υπεύθυνο για τον τρόπο με τον οποίο θα εμφανίζονται τα δεδομένα μας στο διάγραμμα(αλλαγή χρώματος, προσθήκη label και πολλά άλλα). Όλα τα παραπάνω βρίσκονται στο παράθυρο Data.

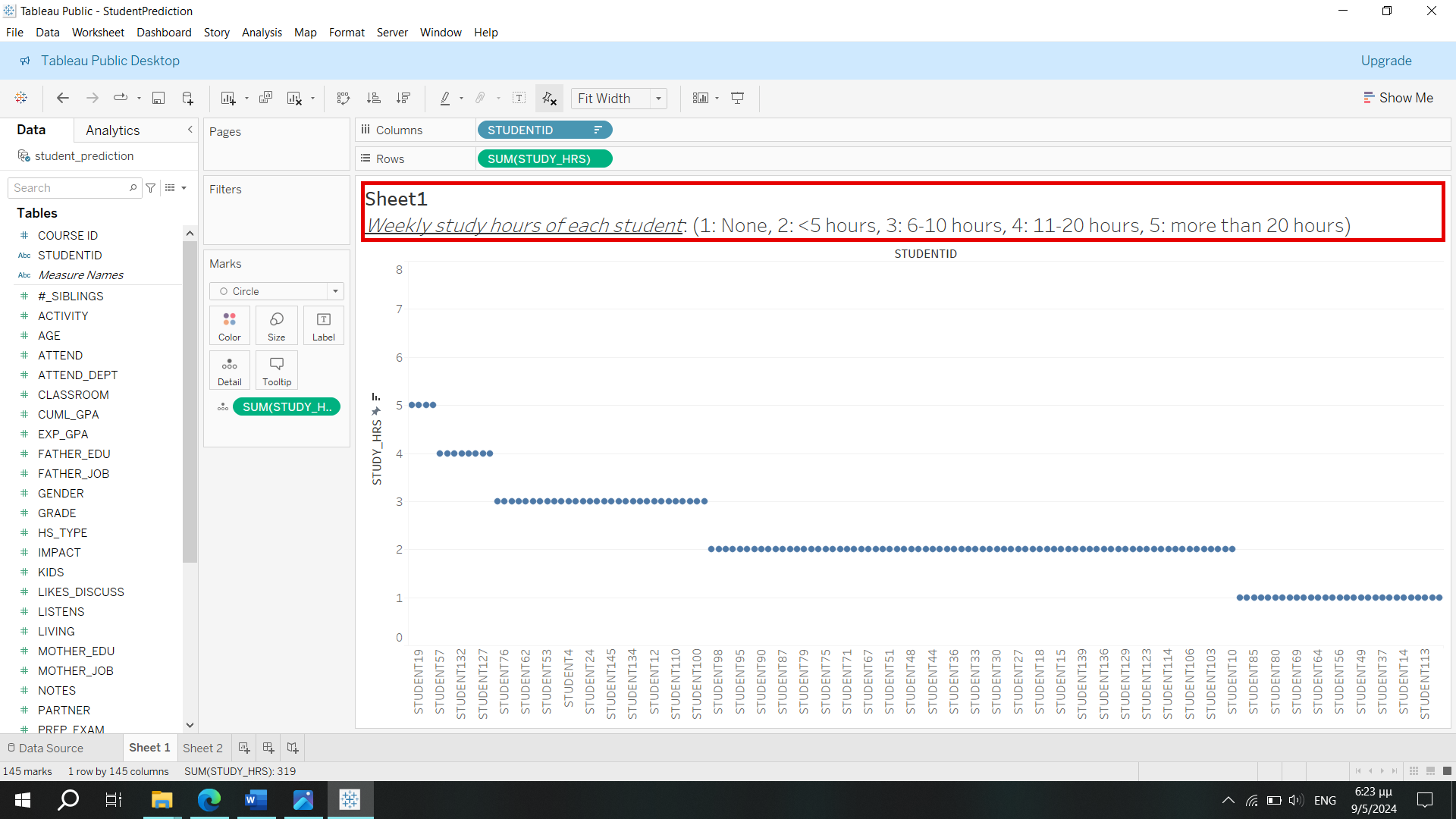
Από την άλλη πλευρά, στο παράθυρο Analytics περιλαμβάνουν διάφορα εργαλεία που μας βοηθούν να αναλύσουμε τα δεδομένα του συνόλου μας και να καταλήξουμε σε ενδιαφέροντα συμπεράσματα. Ειδικότερα, μας δίνεται η δυνατότητα δημιουργίας διάφορων κατανομών, πραγματοποίησης προβλέψεων(χρειάζεται να υπάρχει ημερομηνία ως χαρακτηριστικό στο dataset , για να δούμε πως πιθανώς να διαμορφωθούν στο μέλλον οι τιμές κάποια μεταβλητής), υπολογισμού συναρτήσεων και η δυνατότητα της ομαδοποίησης(clustering) των στιγμιότυπων που εμφανίζονται στο διάγραμμα έχοντας ως μέτρο κάποιο χαρακτηριστικό.

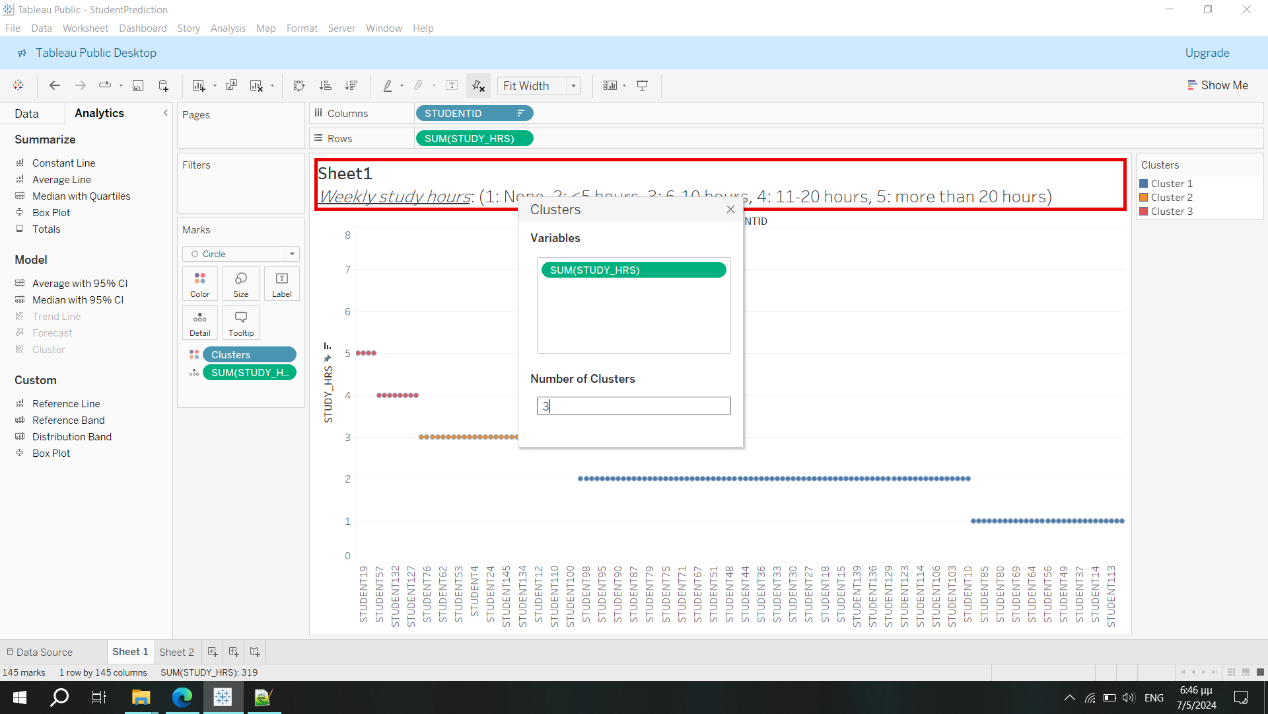
Στο θεωρητικό μέρος της εργασίας αναλύσαμε επαρκώς τη μέθοδο Εξόρυξης Δεδομένων που ονομάζεται ομαδοποίηση και τονίσαμε τους λόγους για τους οποίους είναι τόσο δημοφιλής στον τομέα της ανάλυσης δεδομένων. Οπότε τη χρησιμοποιήσαμε στο παράδειγμα μας για να δείξουμε πως ομαδοποιούνται σε ομάδες οι 145 φοιτητές ανάλογα με τις ώρες που διαβάζουν ανά εβδομάδα. Αναλυτικότερα, το Tableau χρησιμοποιεί την k-means ομαδοποίηση(η ανάλυση της βρίσκεται στο θεωρητικό μέρος), και επιλέξαμε 3 ομάδες, δηλαδή στο Cluster1 βρίσκονται όσοι διαβάζουν λίγες ώρες, στο Cluster3 όσοι διαβάζουν πολλές ώρες και στο Cluster2 όσοι δεν ανήκουν σε καμία από τις προηγούμενες 2 κατηγορίες. Παρουσιάζουμε τα clusters με διάφορους τρόπους(κυκλάκια, bars).

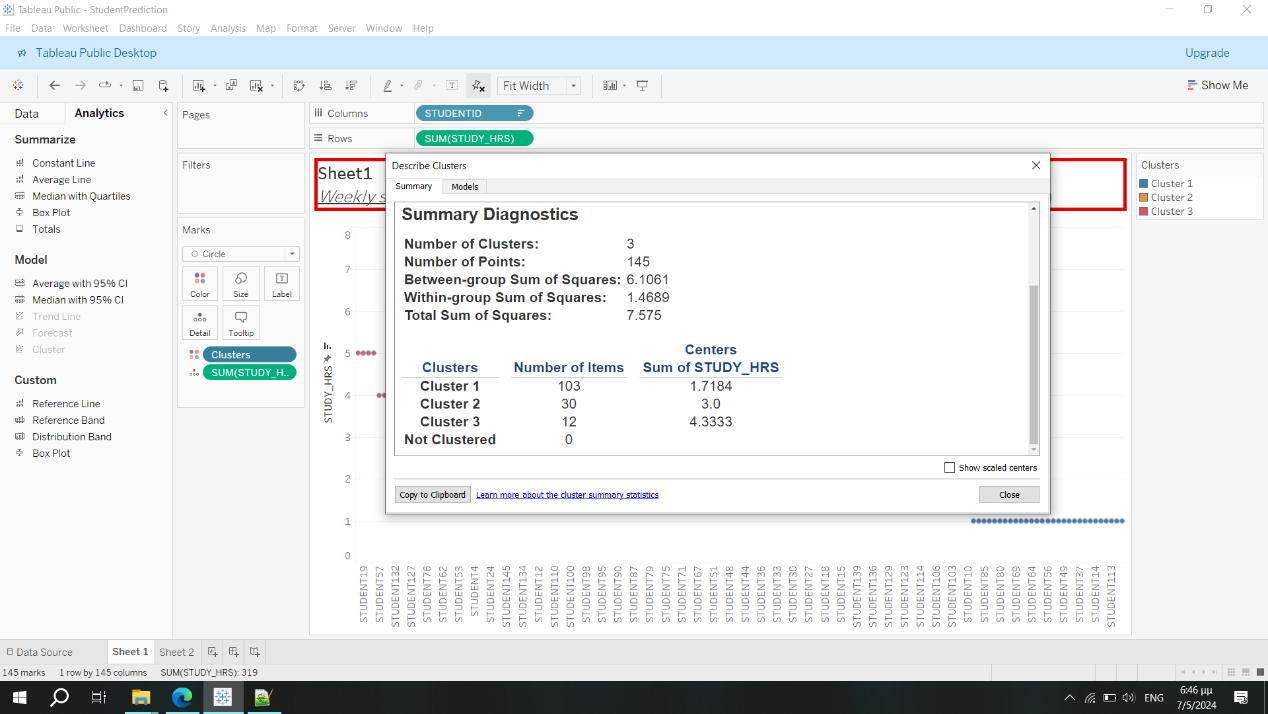
Ακολουθούν όσα περιγράψαμε σε μορφή screenshots μέσα από το περιβάλλον του εργαλείου Tableau. Τα αρχεία του Tableau έχουν κατάληξη .twb, καθώς αποτελούν workbook.

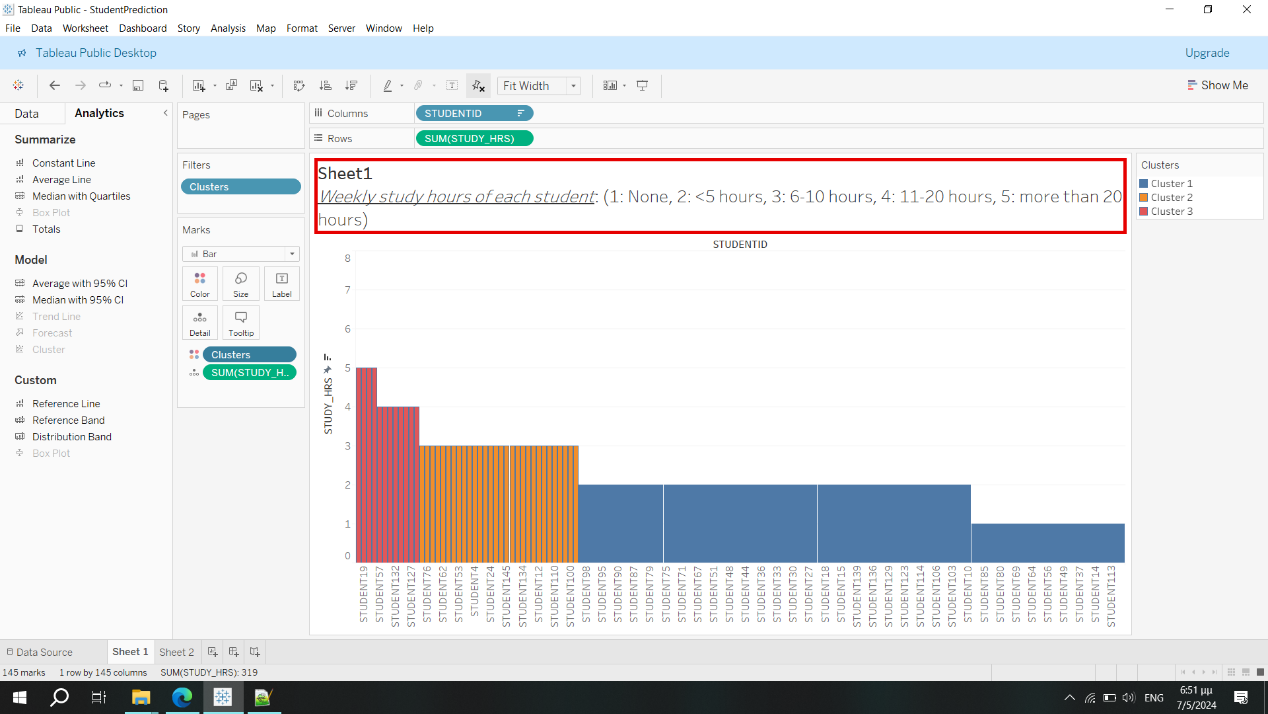


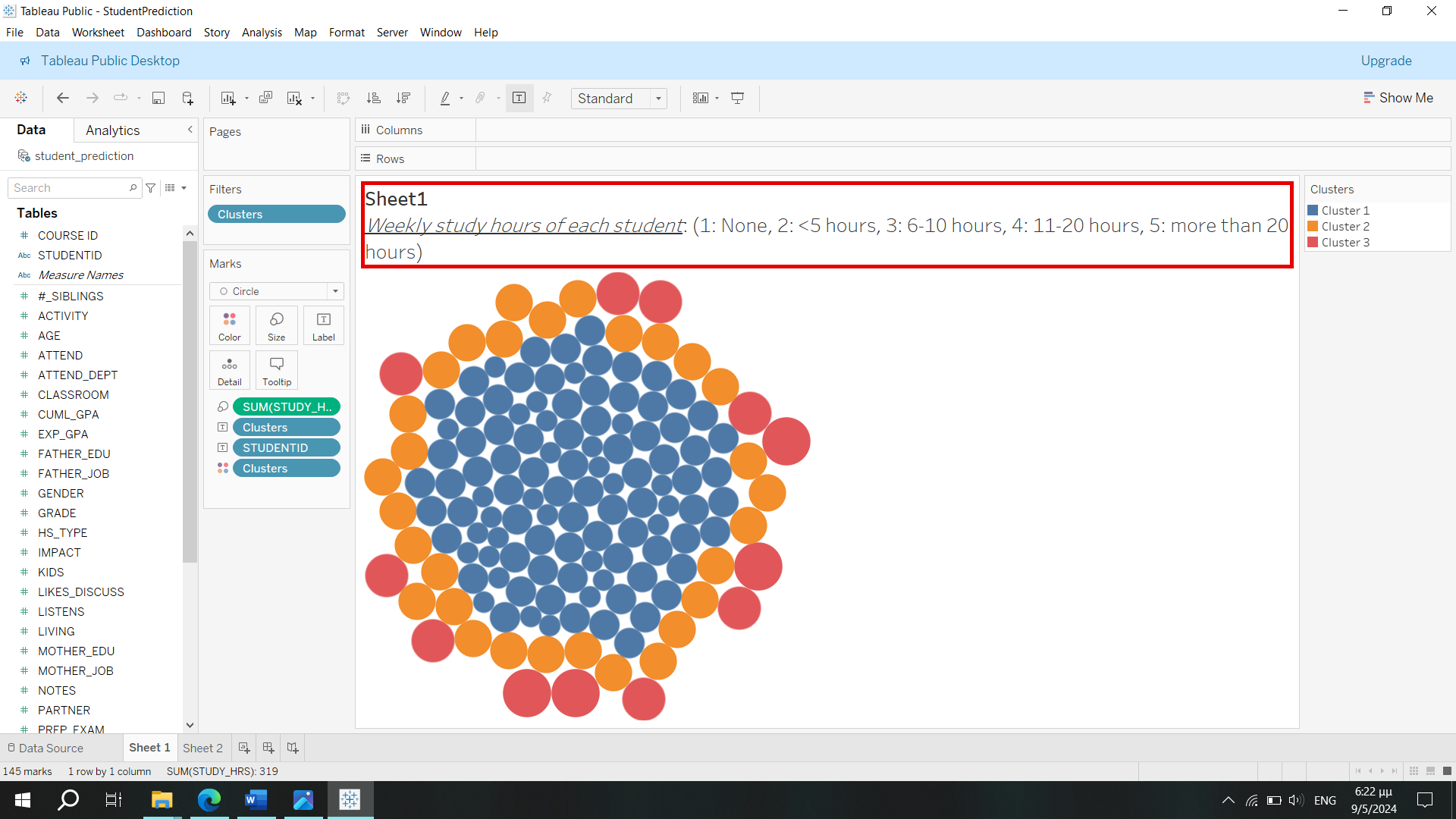






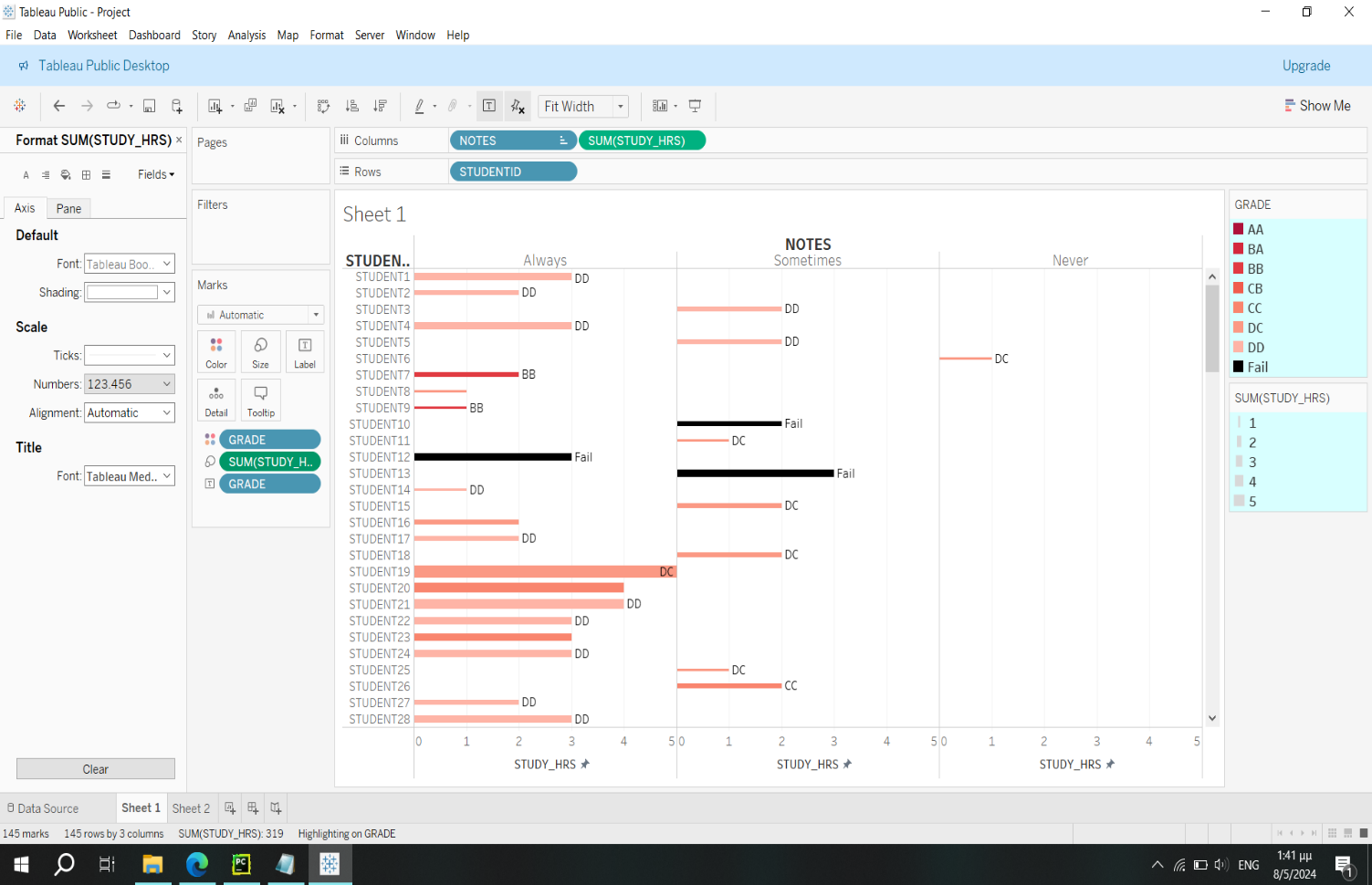






Παράδειγμα 2

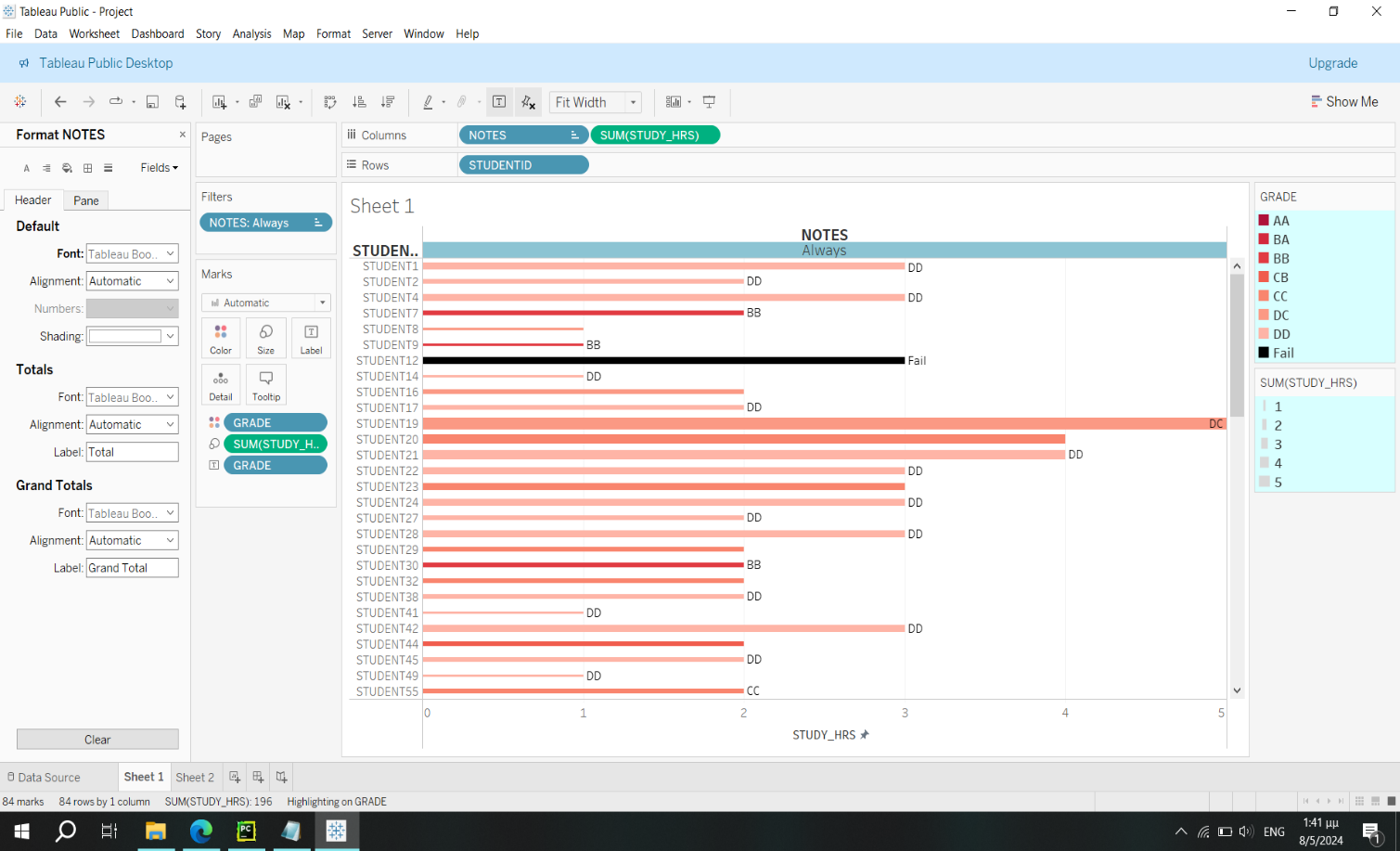
Σε αυτό το σημείο θα αναλύσουμε τους προβλεπόμενους βαθμούς του κάθε φοιτητή σε ένα συγκεκριμένο μάθημα, με βάση τις ώρες που διαβάζει κάθε εβδομάδα και το πόσο συχνά σημειώνει όσα παρουσιάζει στις διαλέξεις ο καθηγητής του μαθήματος. Χρησιμοποιήσαμε το updated dataset, που προήλθε από το αρχείο changes.py που αναφέραμε παραπάνω, ώστε να περιέχουν κατηγορικές τιμές τα χαρακτηριστικά GRADE και NOTES που χρησιμοποιούμε στο συγκεκριμένο παράδειγμα. Στον άξονα x βρίσκονται τα χαρακτηριστικά **NOTES** και **STUDY\_HRS**, άρα για κάθε μία από τις 3 κατηγορικές τιμές(το διάγραμμα χωρίζεται στα 3) που μπορεί να λάβει το NOTES(Always,Sometimes,Never) σημειώνονται και οι ώρες που διαβάζει ο φοιτητής. Στο διάγραμμα, όσο μεγαλύτερη είναι η μπάρα, τόσο περισσότερες ώρες αφιερώνει ο εκάστοτε φοιτητής για να διαβάσει εβδομαδιαίως. Από την άλλη πλευρά, στον άξονα y υπάρχει το χαρακτηριστικό **STUDENTID**. Ενώ, στα Marks τοποθετήσαμε του προβλεπόμενους βαθμούς(**GRADE**) του κάθε φοιτητή για το συγκεκριμένο μάθημα(Βαθμός:Fail). Όσο πιο βαθύ είναι το κόκκινο, τόσο μεγαλύτερος είναι και ο βαθμός του, όμως αν χρωματίστηκε με μαύρο τότε πιθανώς δε θα λάβει προβιβάσιμο βαθμό(υπάρχουν στο δεξί μέρος της εικόνας μας στο περιβάλλον του Tableau).

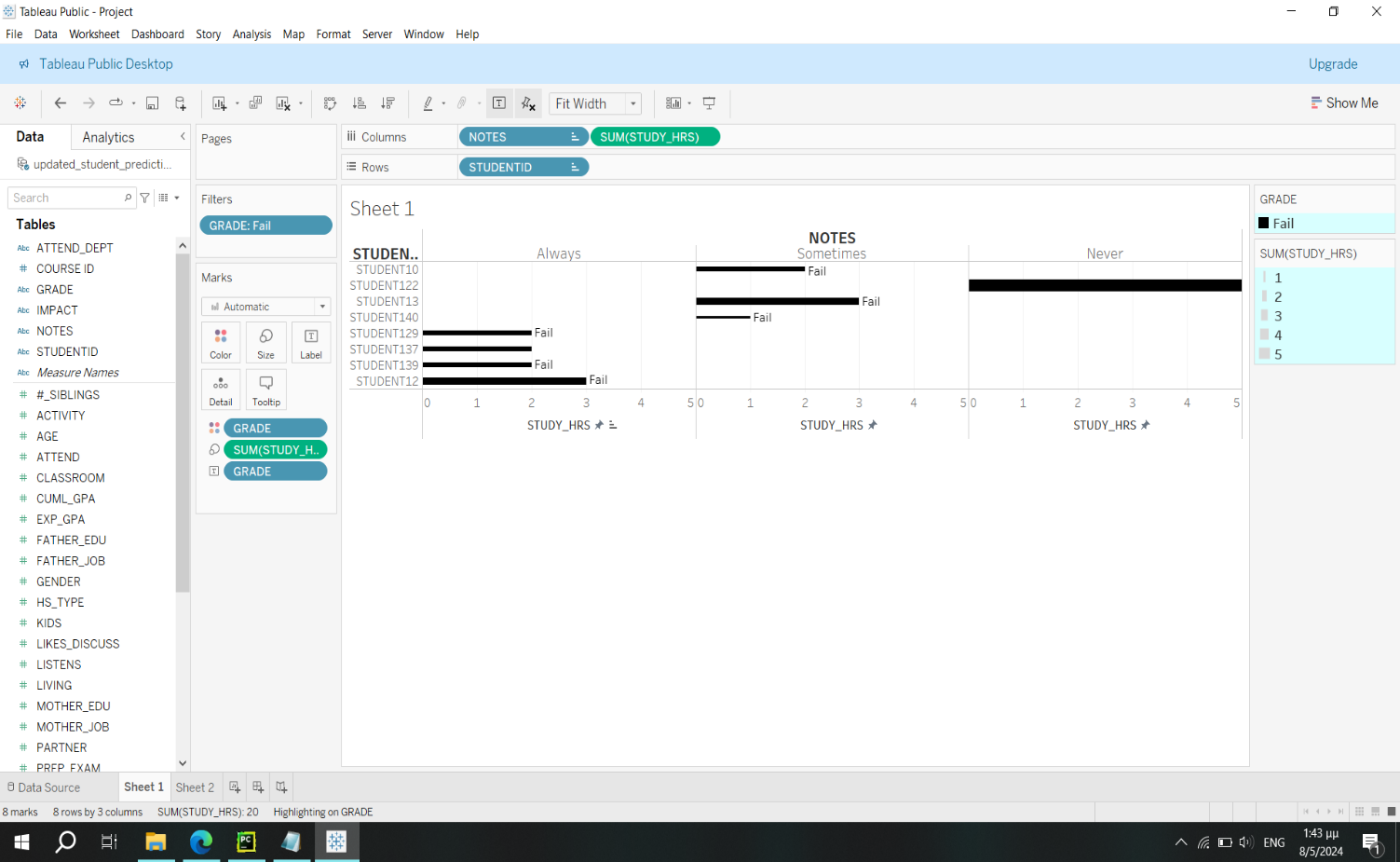


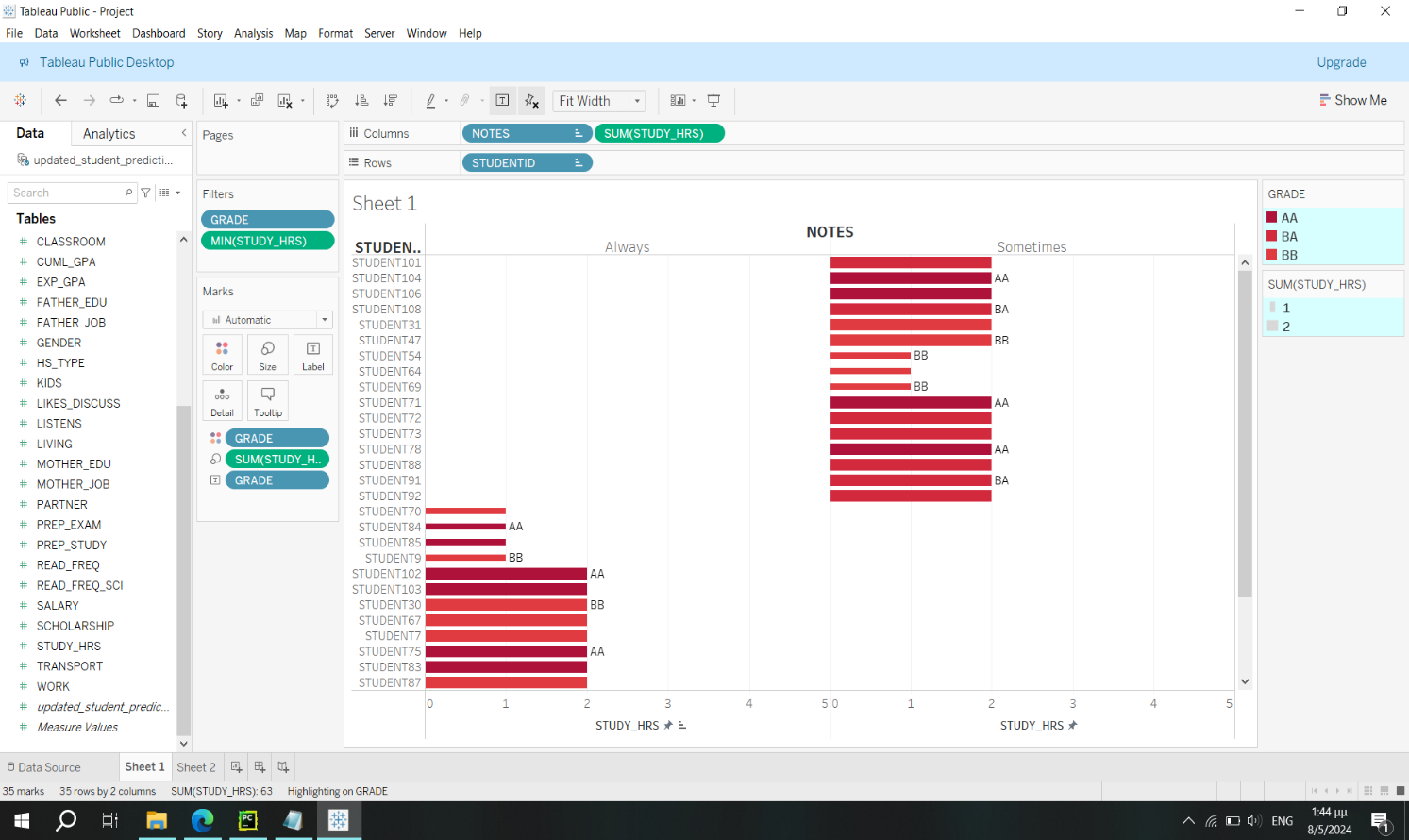
Στη συνέχεια ακολουθούν 3 περιπτώσεις, καθώς πειράζουμε τα φίλτρα ώστε να εξετάσουμε συγκεκριμένες κατηγορίες φοιτητών:

1. Φοιτητές που σημειώνουν πάντα την ώρα του μαθήματος(NOTES:Always). Χρειάζεται ερμηνεία η αποτυχία φοιτητών που ασχολούνται τακτικά με τα μαθήματα τους, διαβάζουν πολλές ώρες, ωστόσο δυσκολεύονται. Είναι χρήσιμο και για τον καθηγητή να συμπεράνει αν οι σημειώσεις που παρέχει στους φοιτητές του τους βοηθούν στις εξετάσεις και σε εργασίες που υποβάλλονται.
2. Φοιτητές με βαθμό Fail(GRADE:Fail). Είναι πολύ σημαντικό να ασχοληθούν καθηγητές, και γενικά οι φορείς που υπάρχουν σε ένα πανεπιστήμιο για να ερμηνεύσουν τους λόγους αποτυχίας των φοιτητών(εκπαιδευτικοί ή κοινωνικοί παράγοντες), ειδικά αν τα ποσοστά αποτυχίας κυμαίνονται σε υψηλά επίπεδα, και να προσπαθήσουν να τους βοηθήσουν.
3. Φοιτητές που διαβάζουν λίγες ώρες(θα τοποθετούνταν στο Cluster1 του Παραδείγματος 1) και ο βαθμός τους είναι πολύ υψηλός(AA/BA/BB). Μέσω αυτής της ανάλυσης θέλουμε να ανακαλύψουμε εάν οι βαθμοί ορισμένων φοιτητών που δε διαβάζουν πολύ έχουν προκύψει από τύχη ή απάτη/αντιγραφή(η ανακάλυψη απάτης στην Εξόρυξη Δεδομένων συνδέεται άρρηκτα με την ανάλυση ακραίων τιμών).

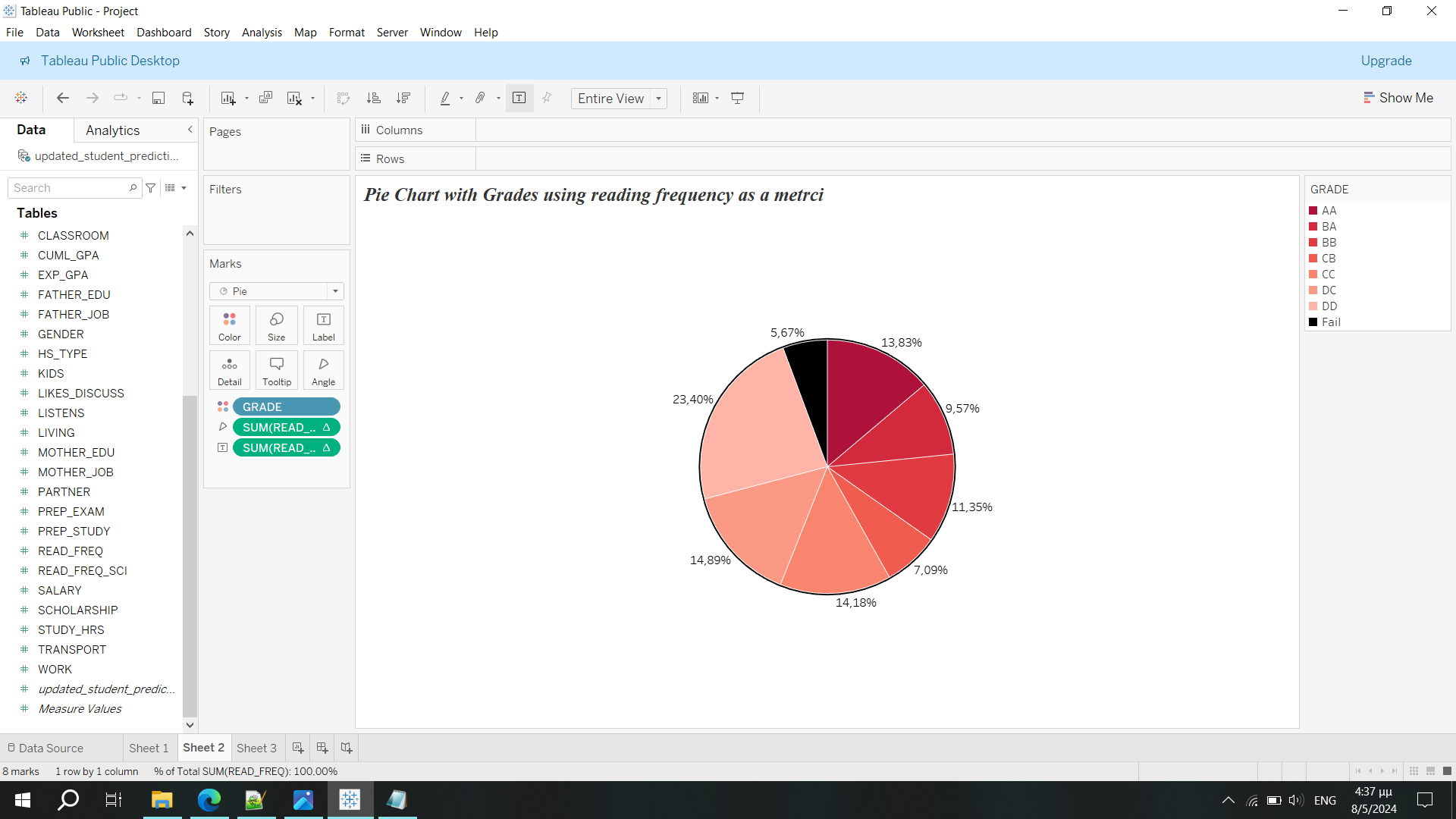
Καταλαβαίνουμε ότι με τα υπο-παραδείγματα που ακολουθούν ίσως ανακαλύψουμε κρυφή γνώση που είναι αρκετά ενδιαφέρουσα. Είναι πιο εύκολο να καταλήξουμε σε συμπεράσματα-αποτελέσματα αν εστιάσουμε σε συγκεκριμένες κατηγορίες φοιτητών, που εμφανίζουν αρκετή ομοιότητα.



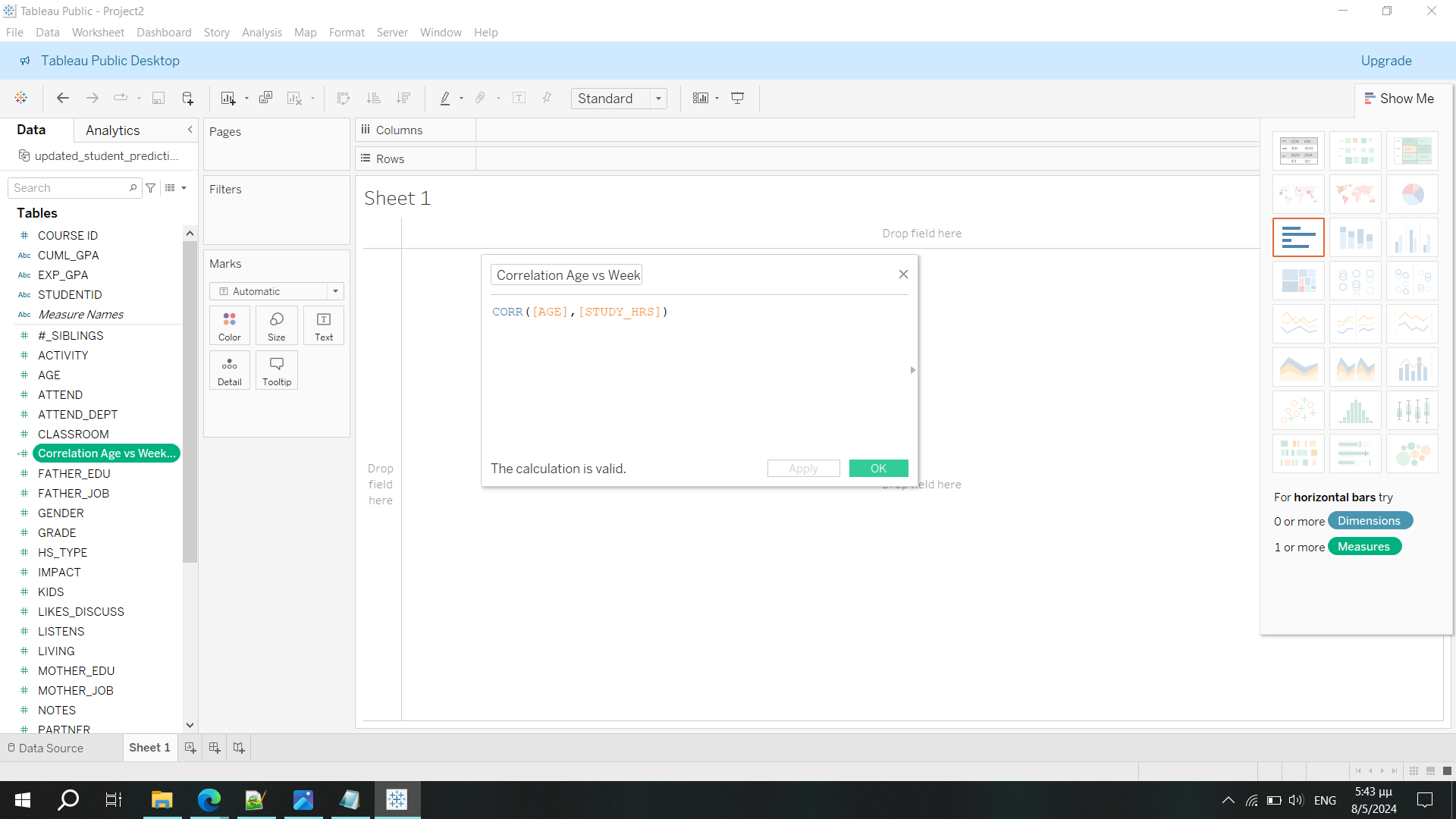


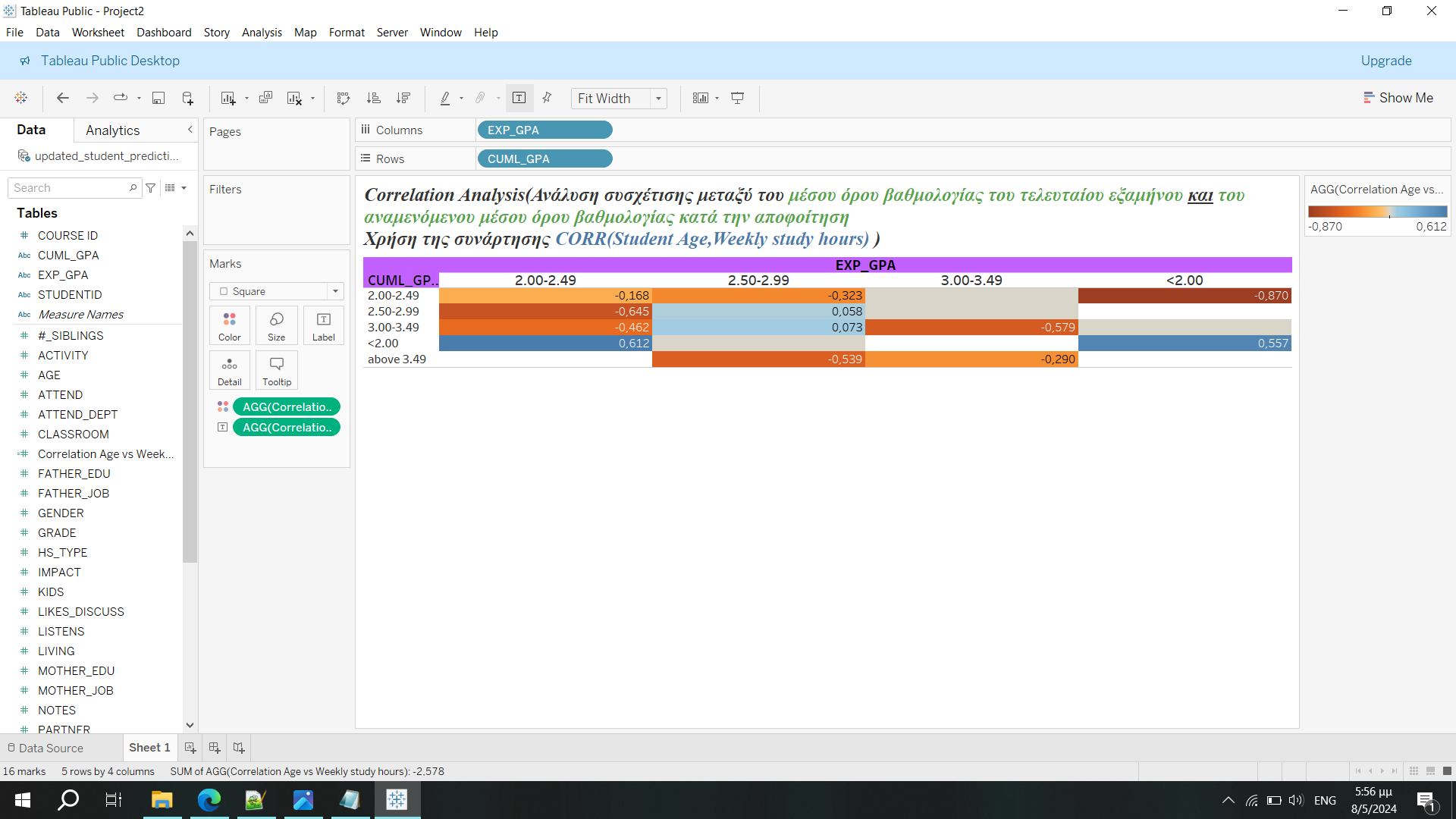


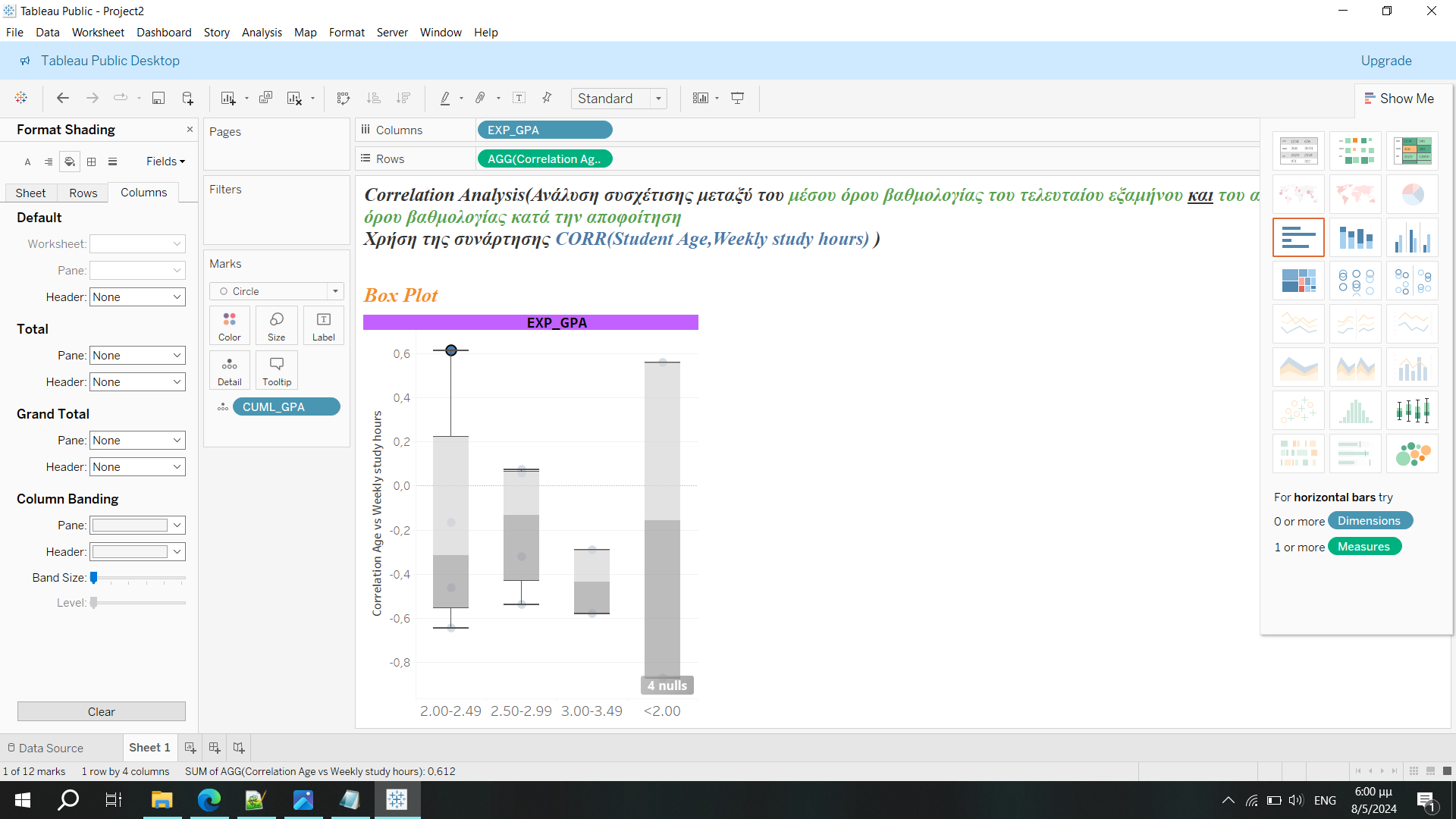
Τέλος, θέλαμε να παρουσιάσουμε ένα διάγραμμα pie. Είναι αρκετά δημοφιλή και μας βοηθούν να καταλάβουμε πως διαμοιράζονται οι βαθμοί των φοιτητών(**GRADE**) αν χρησιμοποιήσουμε τη συχνότητα διαβάσματος βιβλίων(**READ\_FREQ**) ως μετρική. Αναμφίβολα βρισκόμαστε στην περίοδο της Ψηφιακής Επανάστασης, ωστόσο τα βιβλία(χωρίς να σχετίζονται απαραίτητα με το επιστημονικό πεδίο στο οποίο βρίσκεται γη σχολή τους) πάντα θα είναι χρήσιμα σε πολλούς τομείς για τους φοιτητές.



Παράδειγμα 3







Σε αυτό το παράδειγμα ασχολούμαστε με την Ανάλυση Συσχέτισης μεταξύ των ωρών που διαβάζει ένας φοιτητής και της ηλικίας του, με βάση το μέσο όρο της βαθμολογίας που σημείωσε το τελευταίο εξάμηνο και του αναμενόμενου μέσου όρου της βαθμολογίας κατά την αποφοίτηση του(οι τιμές και των 2 χαρακτηριστικών στο συγκεκριμένο πανεπιστήμιο φτάνουν μέχρι το 4). Αποτελεί μέτρο ομοιότητας και ο πιο δημοφιλής συντελεστής συσχέτισης είναι ο Pearson. Παίρνει τιμές από -1 έως 1 και όσο πιο κοντά στο 1 είναι τόσο περισσότερα συσχετισμένα είναι τα χαρακτηριστικά. Μπορεί να χρησιμοποιηθεί σε ένα dataset που περιέχει πολλά χαρακτηριστικά, ώστε να δούμε ποια είναι αυτά που πραγματικά χρειάζονται για την ανάλυση μας. Αν ο συντελεστής συσχέτισης δύο χαρακτηριστικών είναι κοντά στο 1, τότε μπορούμε να κρατήσουμε μόνο το 1 από τα 2 χαρακτηριστικά στο σύνολο δεδομένων και δε θα χάσουμε σημαντική πληροφορία.

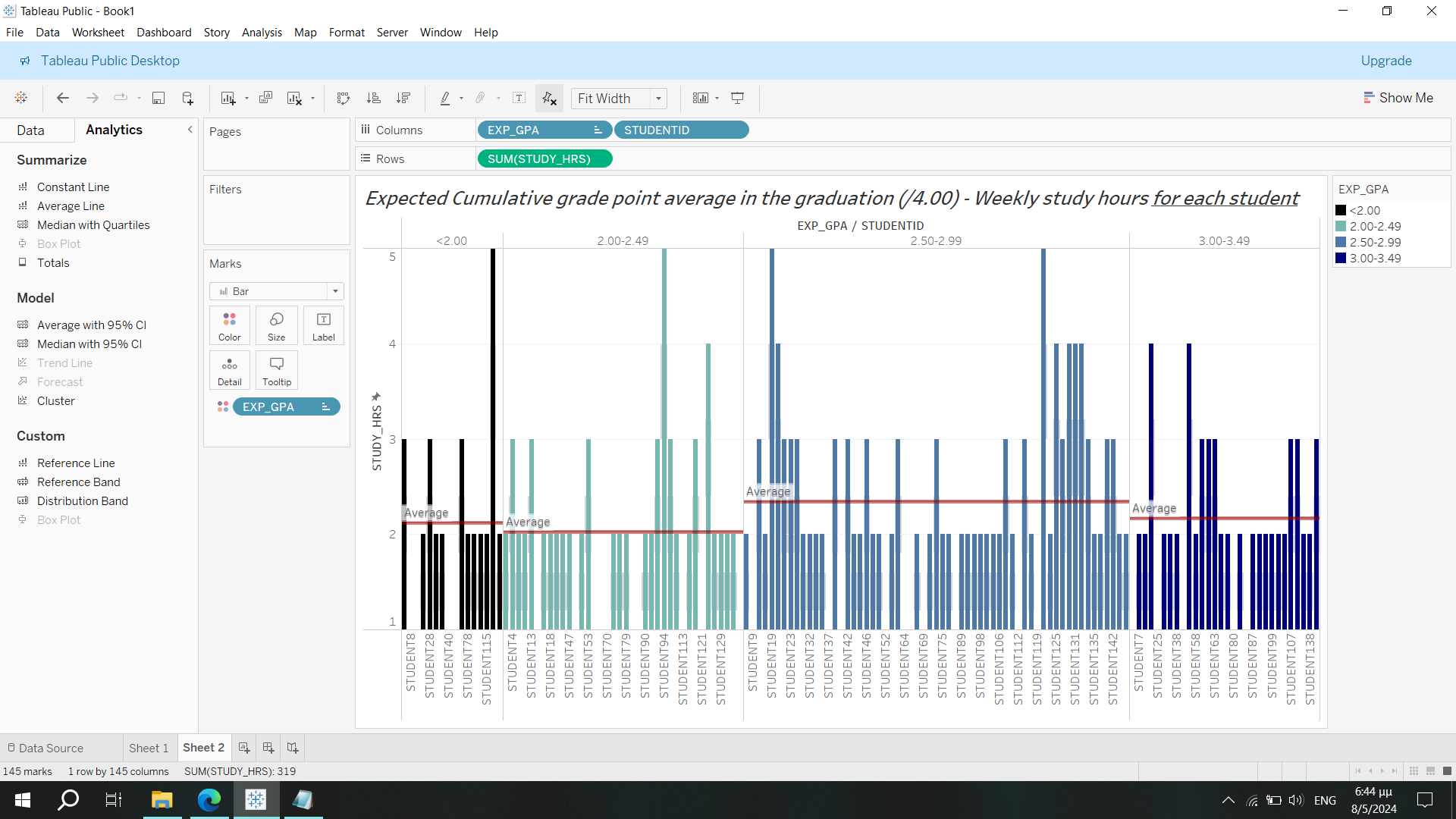
Χρησιμοποιήσαμε τη συνάρτηση CORR([**AGE**],[**STUDY\_HRS**]) για να δημιουργήσουμε το συντελεστή συσχέτισης, και την τοποθετήσαμε στο πεδίο Marks. Έπειτα βάλαμε στους άξονες τα **EXP\_GPA** και **CUML\_GPA**.

Το πρώτο διάγραμμα αποτελείται από τον Πίνακα Συσχέτισης. Στο δεξί μέρος της εικόνας φαίνεται η διακύμανση της συσχέτισης, ανάλογα με το χρώμα του κάθε κελιού του πίνακα. Παρατηρούμε ότι υπάρχει μεγάλη συσχέτιση της ηλικίας του φοιτητή και των ωρών που διαβάζει όταν αναφερόμαστε σε φοιτητές που έχουν χαμηλούς βαθμούς στο τελευταίο εξάμηνο και εκτιμάται ότι και ο βαθμός πτυχίου τους θα είναι χαμηλός.

Όσον αφορά το δεύτερο διάγραμμα, απλώς επιλέξαμε μία από τις προτεινόμενες μορφές διαγραμμάτων, και συγκεκριμένα το box plot.

Γενικά, το συγκεκριμένο παράδειγμα είναι αρκετά πολύπλοκο, αλλά οι συσχετίσεις χρησιμοποιούνται συχνά στην ανάλυση δεδομένων.

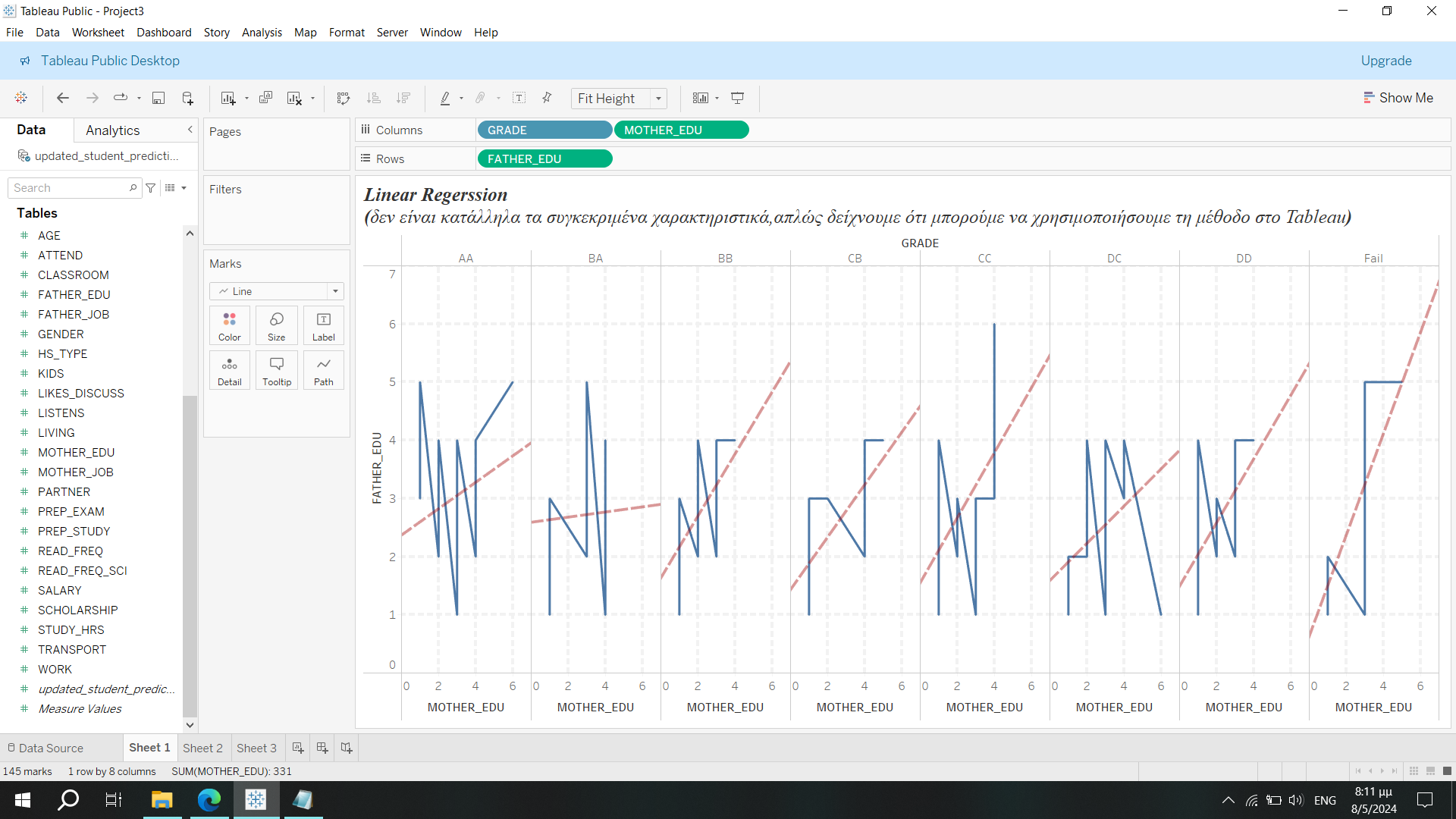
Παράδειγμα 4



Στο συγκεκριμένο διάγραμμα οπτικοποιούμε για κάθε φοιτητή πως είναι πιθανό ότι θα κινηθεί ο βαθμός του(ανάλογα με το χρώμα που φαίνεται στο δεξί μέρος της εικόνας) σε σύγκριση με τον αριθμό των ωρών που διαβάζει σε εβδομαδιαία βάση(ανάλογα με το ύψος της μπάρας). Επίσης μέσω του Average Line που βρίσκεται στο παράθυρο Analytics παρουσιάζουμε το μέσο όρο των τιμών με κόκκινη γραμμή.

Εντύπωση προκαλεί το γεγονός ότι ο φοιτητής 115 αναμένεται να έχει χαμηλό βαθμό στην αποφοίτηση, ενώ διαβάζει πάρα πολλές ώρες. Αντίθετα, υπάρχουν αρκετοί συμφοιτητές του που είναι πιθανό να πάρουν υψηλό βαθμό πτυχίου χωρίς να διαβάζουν πολλές ώρες. Τέτοιες παρατηρήσεις πρέπει να αναλυθούν σε περισσότερο βάθος και να ζητηθεί η γνώμη ειδικών σε εκπαιδευτικά θέματα.

Παράδειγμα 5

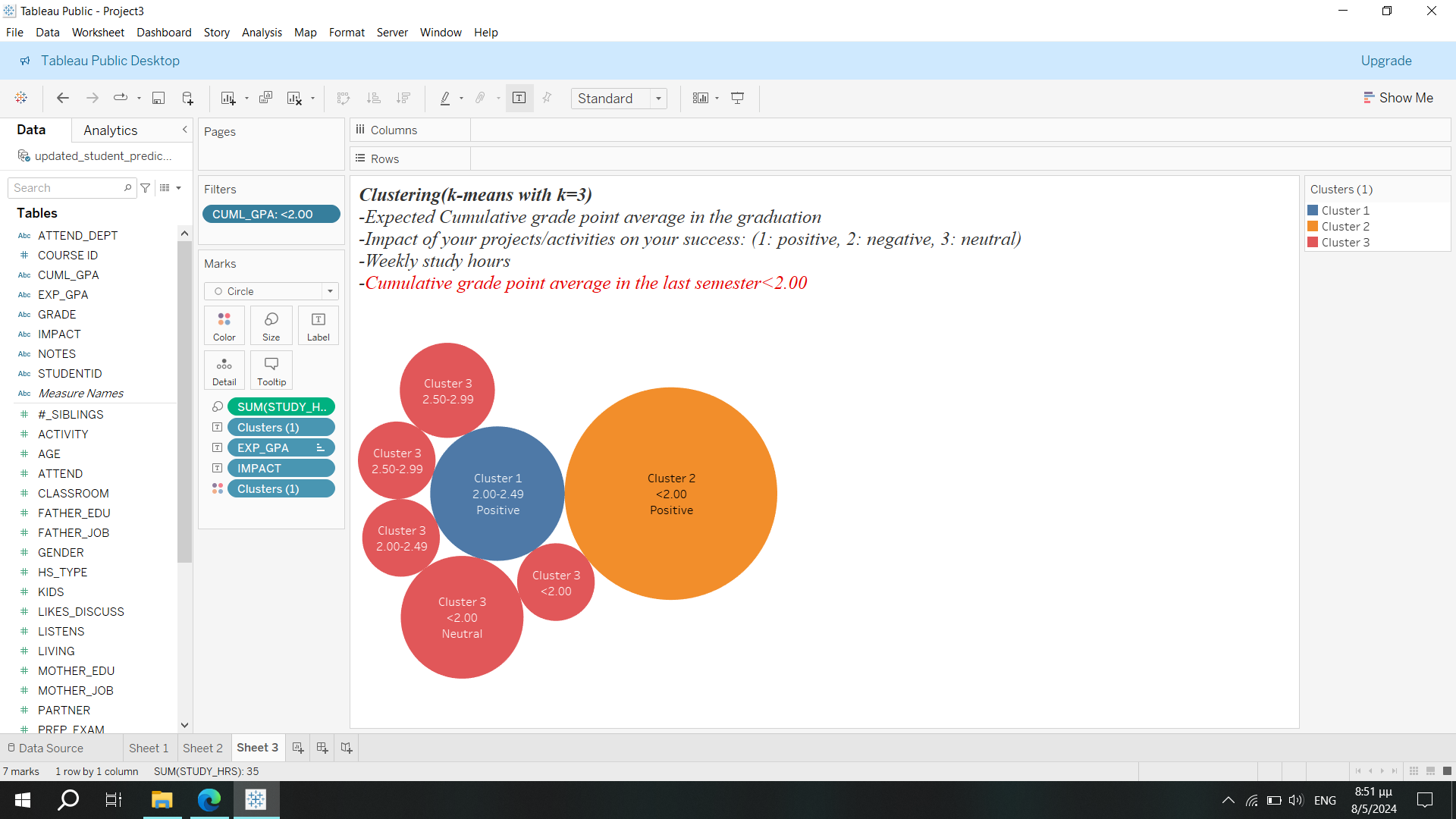


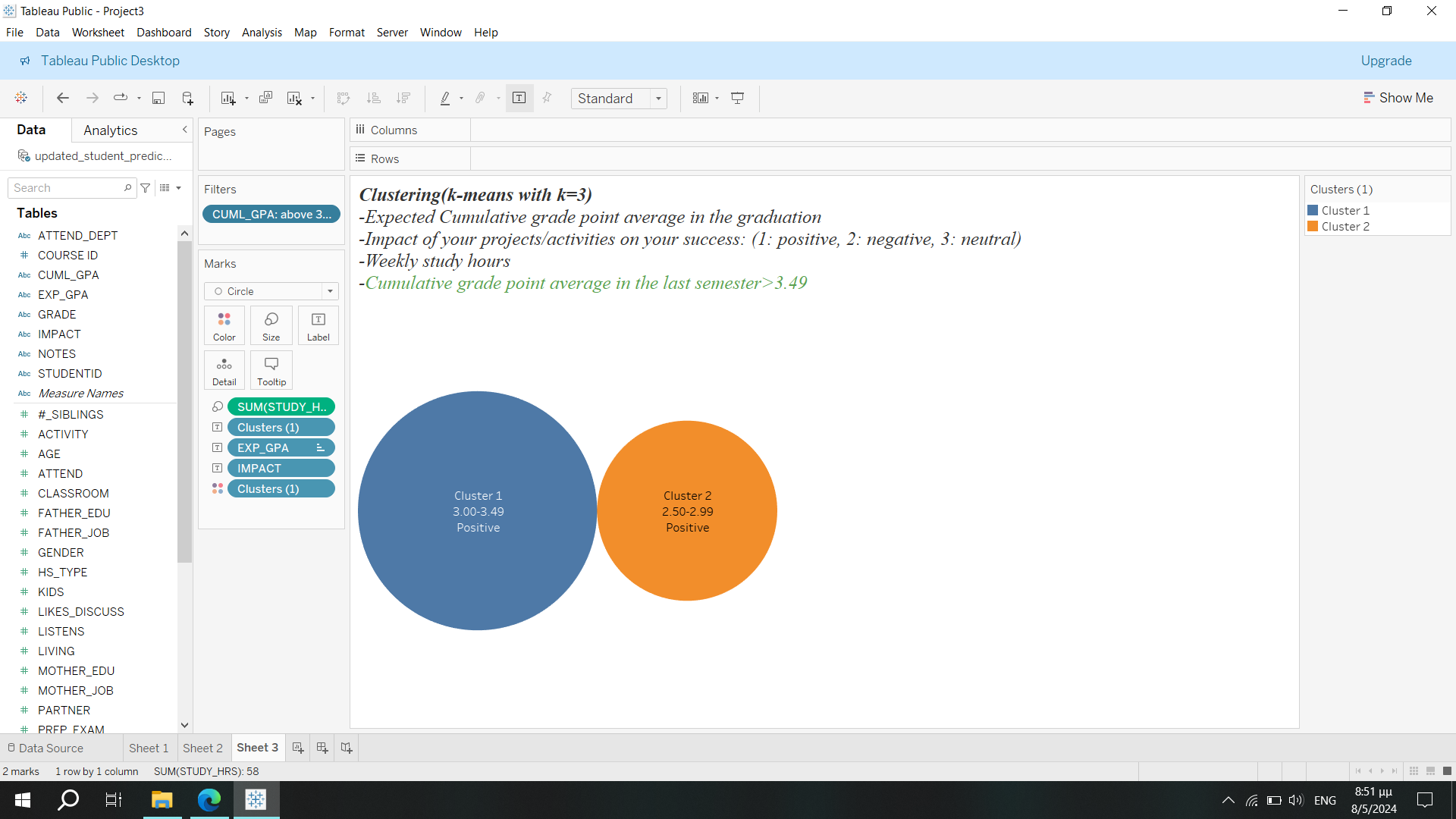
Στην παραπάνω εικόνα παρουσιάζεται η εφαρμογή της Γραμμικής Παλινδρόμησης. Γενικά η Παλινδρόμηση αποτελεί μία αρκετά χρήσιμη μέθοδο Εξόρυξης Δεδομένων και αναλύθηκε στο Θεωρητικό μέρος της εργασίας. Τα χαρακτηριστικά του συνόλου δεδομένων που επιλέξαμε δεν είναι κατάλληλα για την επίδειξη της χρήσης της Γραμμικής Παλινδρόμησης. Ωστόσο, θέλαμε να δείξουμε ότι μπορεί να δημιουργηθεί στο εργαλείο Tableau και προσπαθήσαμε να φτιάξουμε ένα σενάριο που έχει και κάποια κοινωνική «σφραγίδα». Ειδικότερα, επιλέξαμε στους άξονες του διαγράμματος μας τα χαρακτηριστικά που αφορούν το επίπεδο εκπαίδευσης της μητέρας και του πατέρα του κάθε φοιτητή(**MOTHER\_EDU** , **FATHER\_EDU**), και χωρίσαμε το διάγραμμα σε 8 κομμάτια, κάθε ένα από τα οποία αποτελεί έναν βαθμό(**GRADE**).

Αναμφίβολα υπάρχουν διαφοροποιήσεις μεταξύ του επιπέδου εκπαίδευσης των δύο γονέων ενός φοιτητή, αλλά και μεταξύ του επιπέδου εκπαίδευσης των γονέων και της επίδοσης του φοιτητή. Κοινωνικοί παράγοντες ενδέχεται να επηρεάζουν την απόδοση και τις ευκαιρίες που έχουν οι φοιτητές όσον αφορά τις σπουδές του, ωστόσο είναι ένα ιδιαίτερο και αρκετά πολύπλοκο ζήτημα για ανάλυση.

Οπότε, εξετάζουμε τη σχέση μεταξύ του βαθμού των φοιτητών και του επιπέδου εκπαίδευσης των γονέων τους. Προσπαθούμε δηλαδή να προβλέψουμε το βαθμό του φοιτητή χρησιμοποιώντας δύο χαρακτηριστικά του συνόλου δεδομένων. Δηλαδή, με τη χρήση της μεθόδου της γραμμικής παλινδρόμησης αλλά και των υπόλοιπων μεθόδων πρόβλεψης έχουμε τη δυνατότητα να προβλέψουμε πως θα κινηθεί μία μεταβλητή, στη συγκεκριμένη περίπτωση ο βαθμός του φοιτητή.

Παράδειγμα 6





Στο τελευταίο παράδειγμα του Πρακτικού Μέρους της εργασίας χρησιμοποιούμε άλλη μία φορά τη μέθοδο της Ομαδοποίησης(Clustering). Επιλέξαμε 3 χαρακτηριστικά από το σύνολο δεδομένων: Τον εκτιμώμενο βαθμό πτυχίου, το αντίκτυπο των project στα οποία συμμετέχει και έχει παραδώσει ο φοιτητής κατά τη διάρκεια του εξαμήνου(**IMPACT**) και τον αριθμό των ωρών που διαβάζει κάθε εβδομάδα. Επίσης χρησιμοποιούμε ως φίλτρο το συνολικό βαθμό του φοιτητή στο προηγούμενο εξάμηνο, και με βάση αυτό το μέτρο δημιουργούνται συγκεκριμένες ομάδες αυτόματα από το Tableau, που εφαρμόζει k-means ομαδοποίηση. Αξίζει να σημειωθεί ότι ο τίτλος του κάθε διαγράμματος που έχουμε προσθέσει βοηθάει στην κατανόηση του γραφήματος που παρουσιάζεται.

Αναλυτικότερα, στην πρώτη εικόνα παρουσιάζονται οι φοιτητές που είχαν συνολικό βαθμό<2 στο προηγούμενο εξάμηνο και με βάση το μέτρο ομοιότητας που αναφέραμε παραπάνω χωρίζονται σε 3 ομάδες. Από την άλλη πλευρά, στη δεύτερη εικόνα δημιουργήθηκαν 2 ομάδες στις οποίες συμμετέχουν φοιτητές με συνολικό βαθμό>3.49(τιμή «ταβάνι» είναι το 4) στο προηγούμενο εξάμηνο, χωρίς να αλλάξουμε κάτι άλλο σε σχέση με την αμέσως προηγούμενη εικόνα.

Άρα, θέλουμε να εξετάσουμε δύο κατηγορίες, τους φοιτητές που είχαν χαμηλό συνολικό βαθμό στο προηγούμενο εξάμηνο και αυτούς που είχαν άριστες επιδόσεις. Τα υπόλοιπα τρία χαρακτηριστικά του παραδείγματος είναι ενδεικτικά, και θα μπορούσαμε να χρησιμοποιήσουμε κάποιον άλλο συνδυασμό χαρακτηριστικών, ανάλογα με το πρόβλημα που θέλουμε να αναδείξουμε και εξαρτάται από τον αποδέκτη της ανάλυσης δεδομένων(καθηγητής ,φοιτητής, Υπουργείο Παιδείας).