Βασίλειος Δημητρίου - Capstone

**Εγώ ο Βασίλειος Δημητρίου βεβαιώνω ότι η παρούσα εργασία βασίζεται σε δεδομένα για τα οποία έχω άδεια μεταβίβασης και επεξεργασίας. Βεβαιώνω επίσης ότι τα δεδομένα αυτά τηρούν όλους τους ισχύοντες νόμους και διατάξεις περί προστασίας πνευματικής ιδιοκτησίας και προσωπικών δεδομένων.**

Η ατομική εργασία είναι αναρτημένη στον ακόλουθο σύνδεσμο, όπου περιλαμβάνονται και τα δεδομένα που χρησιμοποίησα αλλά και η ανάλυση και τα αποτελέσματα.

[Vasilisdi/Data-Analyst-NTUA-Course---Vasileios-Dimitriou: Assignment in the context of the NTUA course (github.com)](https://github.com/Vasilisdi/Data-Analyst-NTUA-Course---Vasileios-Dimitriou)

## ΕΠΙΛΟΓΗ / ΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ-(l2036-c1-u1)

Στα πλαίσια της εξέτασης στο μάθημα επιλογή/επεξεργασία δεδομένων (i2036-c1-u1), γίνεται χρήση του repository Breast Cancer Wisconsin (Diagnostic) Data Set. Πιο συγκεκριμένα από το Data Folder του, θα γίνει αξιοποιηθούν τα αρχεία **«breast-cancer-wisconsin.data»** και **«breast-cancer-wisconsin.names»,** τα οποία περιλαμβάνουν τα πεδία:

* "Sampldate code number",
* "Clump Thickness",
* "Uniformity of Cell Size",
* "Uniformity of Cell Shape" ,
* "Marginal Adhesion" ,
* "Single Epithelial Cell Size" ,
* "Bare Nuclei" ,
* "Bland Chromatin" ,
* "Normal Nucleoli" ,
* "Mitoses " ,
* "Class"

Παρατηρούμε λοιπόν ότι το πεδίο "Bare Nuclei" , περιλαμβάνει κάποια «?»-ερωτηματικά, για τα οποία έχει κρίθηκε ωφέλιμο το να απαλειφθούν. Μη έχοντας κάποια περεταίρω δεδομένα συμπληρώθηκαν στοχαστικά.

## ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ-(l2036-c1-u2)

Όσο μεγαλύτερη η τιμή τόσο πιο πολύ τείνει να εμφανίσει κακοήθεια.