

Μάθημα: ΜΕΘΟΔΟΙ ΣΤΑΤΙΣΤΙΚΗΣ ΚΑΙ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ

Ακαδημαϊκή περίοδος: Χειμερινό εξάμηνο 2022-2023

1^η Εργασία

Το αρχείο cities.csv περιλαμβάνει δεδομένα σχετικά με τον τρόπο ζωής σε 44 πόλεις στον κόσμο κατά το έτος 2021. Συγκεκριμένα, έχουν καταγραφεί οι ακόλουθες μεταβλητές:

city	Πόλη
sunshine	Σύνολο ωρών ηλιοφάνειας
pollution	Δείκτης ρύπανσης
work	Μέσος ετήσιος χρόνος εργασίας (σε ώρες)
activities	Συνολικός αριθμός υπαίθριων δραστηριοτήτων στην πόλη
places	Πλήθος επιλογών για διασκέδαση στην πόλη (καφέ, εστιατόρια, μπαρ κλπ.)
bottle	Κόστος ενός μπουκαλιού νερό (σε λίρες)
membership	Μέσο μηνιαίο κόστος συνδρομής σε γυμναστήριο (σε λίρες)

Στόχος είναι οι πόλεις να ομαδοποιηθούν με τρόπο που να αποτυπώνει το επίπεδο ζωής των πολιτών τους. Για το σκοπό αυτό, θα πρέπει να εφαρμόστε διαφορετικές τεχνικές συσταδοποίησης, απαντώντας παράλληλα στα ακόλουθα ερωτήματα:

- i. Ποια μέθοδος είναι προτιμότερη για τα δεδομένα μας και γιατί;
- ii. Πόσο επιτυχημένη μπορεί να θεωρηθεί συνολικά η ομαδοποίηση στην οποία καταλήξατε;
- iii. Υπάρχουν ομάδες που να αναγνωρίζονται με σαφήνεια από όλες τις μεθόδους και ομάδες που να μην είναι και τόσο ευδιάκριτες;
- iv. Ερμηνεύστε τα αποτελέσματα της ομαδοποίησης στην οποία καταλήξατε λαμβάνοντας υπόψη όλη τη διαθέσιμη πληροφορία.

Τα παραδοτέα της εργασίας περιλαμβάνουν:

- Το script file με τον κώδικά σας,
- Μία γραπτή αναφορά που θα περιλαμβάνει **αιτιολογημένες** απαντήσεις στα παραπάνω ερωτήματα και μία παράγραφο που να συνοψίζει τις τελικές σας σκέψεις και συμπεράσματα. Όπως σε κάθε ανάλυση, θα πρέπει προφανώς να ξεκινήσετε και να συμπεριλάβετε στην παρουσίαση των αποτελεσμάτων σας βασικά περιγραφικά μέτρα και γραφήματα για όλες τις μεταβλητές.

Η εργασία θα βαθμολογηθεί με άριστα το 10 και θα μετρήσει κατά 10% στον τελικό σας βαθμό.

Η εργασία θα πρέπει να αναρτηθεί στο eclass μέχρι την **Κυριακή 8 Ιανουαρίου 2023** στις 24:00.
Καμία εργασία δε θα γίνει δεκτή μετά από τη συγκεκριμένη ημερομηνία και ώρα.

Καλή επιτυχία!