

ΜΗΝ ΑΚΟΥΣΩ ΤΙΠΟΤΑ ΓΙΑ ΤΟ ΕΞΩΦΥΛΛΟ

ΤΟ ΞΕΡΩ ΠΩΣ ΕΙΝΑΙ ΣΚΑΤΑ



ΛΙΓΗ ΥΠΟΜΟΝΗ

ΚΑΙ ΘΑ ΤΟ ΦΤΙΑΞΩ

ΣΕ ΑΡΓΟΤΕΡΟ COMMIT

# Εισαγωγή

Η ανάπτυξη της εφαρμογής έγινε με τη γλώσσα Python, με το interface του Jupyter, χρησιμοποιώντας το Visual Studio Code και διάφορες χρήσιμες βιβλιοθήκες, που βοήθησαν πολύ με την οπτικοποίηση και την παλινδρόμηση των δεδομένων. Παρακάτω θα εξηγηθούν αναλυτικά οι χρήσεις των βιβλιοθηκών και η σημασία τους στη διευκόλυνση πολλών διαδικασιών. Συγκεκριμένα, οι βιβλιοθήκες που χρησιμοποιούνται είναι οι:

* numpy
* pandas
* matplotlib
* seaborn
* sklearn

Όλες οι παραπάνω είναι απαραίτητες για την εκτέλεση της εφαρμογής.

# Προ-επεξεργασία Δεδομένων

## ΜΕΡΟΣ 1ο: Αναγνώριση της Πληροφορίας

Το αρχείο CSV, που περιέχει τα δεδομένα, διαβάστηκε με τη χρήση της βιβλιοθήκης pandas, ώστε να έχουμε πλήρη εικόνα των μετρήσεων των δεδομένων. Όπως φαίνεται και παρακάτω, τα δεδομένα που έχουμε είναι 8 (οκτώ) στήλες τύπου αριθμού κινητής υποδιαστολής και μίας στήλης τύπου αλφαριθμητικού.

A screenshot of a computer

Description automatically generated with medium confidence

Εικόνα 1: Τα πρώτα 5 (πέντε) δεδομένα του αρχείου CSV.

Text

Description automatically generated

Εικόνα 2: Αναγνώριση των τύπων δεδομένων των στηλών.

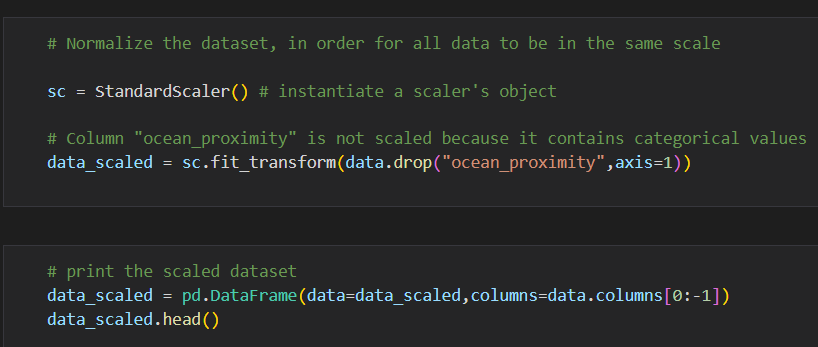
## Μέρος 2ο: Κλιμάκωση Δεδομένων

Εφ’ όσον έχουμε οκτώ στήλες στις οποίες εμπεριέχονται αριθμοί, θα πρέπει να κάνουμε κλιμάκωση. Δηλαδή, θα πρέπει όλες οι τιμές να ανήκουν στο ίδιο διάστημα. Όσο για την άλλη στήλη, που δεν έχει από μόνη της αριθμητικά δεδομένα, θα ακολουθήσουμε άλλη μεθοδολογία, ώστε να έχουμε αριθμητικές τιμές. Παρατηρούμε ότι έχει 5 (πέντε) διαφορετικές τιμές:

* <1H OCEAN
* INLAND
* ISLAND
* NEAR BAY
* NEAR OCEAN

Άρα, θα διαγράψουμε τη στήλη που έχει αλφαριθμητικές τιμές και αντ’ αυτής, θα βάλουμε πέντε επιπλέον στήλες. Αυτές οι πέντε στήλες θα αντιστοιχούν στα παραπάνω ονόματα.

Για αρχή, κλιμακώνουμε τα δεδομένα μας, ώστε όλα να ανήκουν στο σύνολο [-3, 3].



Εικόνα 3: Κλιμάκωση των στηλών με τις αριθμητικές τιμές.

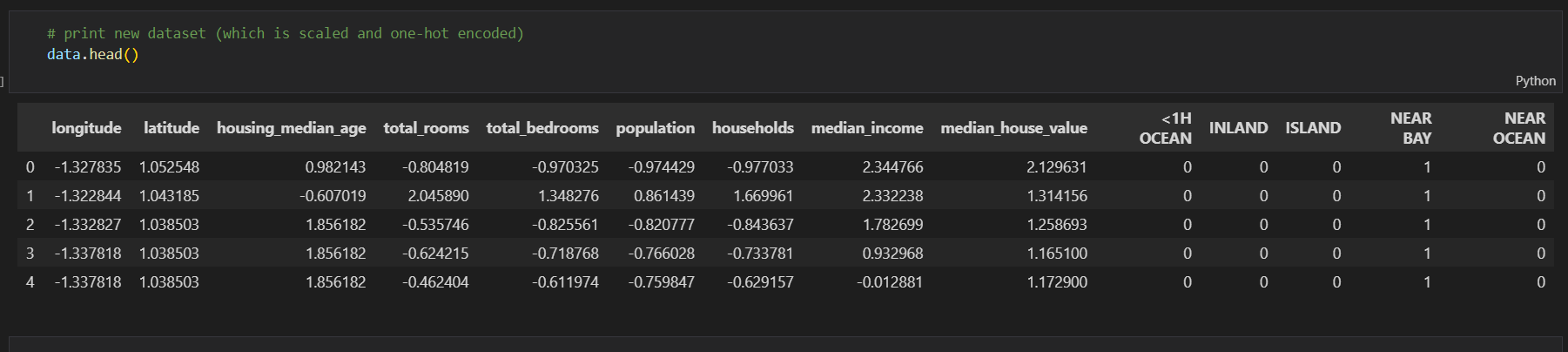
Ύστερα, «μαζεύουμε» τα δεδομένα της στήλης στην οποία εμπεριέχονται τα αλφαριθμητικά και τα κάνουμε όλα πέντε ξεχωριστές στήλες.

Text

Description automatically generated

Εικόνα 4: Συμμάζεμα των δεδομένων της στήλης με τα αλφαριθμητικά.

Οπότε, η τελική μορφή του πίνακα, μετά την κλιμάκωση των δεδομένων, είναι η εξής:



Εικόνα : Τα δεδομένα του πίνακα μετά την κλιμάκωση.