

Αρχές Γλωσσών Προγραμματισμού & Μεταφραστών

Τμήμα Μηχανικών Η/Υ & Πληροφορικής

Πανεπιστήμιο Πατρών

Εαρινό Εξάμηνο 2024

Διδάσκοντες: Ιωάννης Γαροφαλάκης, Σπυρίδων Σιούτας, Παναγιώτης Χατζηδούκας

Προαιρετική Εργαστηριακή Άσκηση Python

Η γλώσσα Python περιλαμβάνει μια σειρά από εξωτερικά πακέτα και βιβλιοθήκες λογισμικού και αποτελεί ένα ισχυρό εργαλείο ανοικτού κώδικα, το οποίο έχει μεγάλη χρησιμότητα εκτός από την αυστηρά οριοθετημένη περιοχή της επιστήμης των υπολογιστών, και στην έρευνα κάθε ειδικότητας αναφορικά με την ανάλυση και την επεξεργασία των δεδομένων.

Είναι στην ουσία μια γλώσσα προγραμματισμού γενικού σκοπού, αντίθετα με άλλες γλώσσες οι οποίες επικεντρώνονται αυστηρά σε συγκεκριμένους τομείς όπως π.χ. η στατιστική ανάλυση. Το μεγάλο της πλεονέκτημα αφορά τη δυνατότητα της συγγραφής κώδικα στον οποίο μπορεί να γίνεται κάθε φορά χρήση των βιβλιοθηκών που είναι σχετικές με τις ζητούμενες επιστημονικές περιοχές του εκάστοτε project.

Στην ιστοσελίδα <https://www.kaggle.com/datasets/mojtaba142/hotel-booking> θα βρείτε το dataset "Hotel Booking" που περιέχει πληροφορίες για κρατήσεις σε ένα City Hotel και ένα Resort Hotel, με διάφορα χαρακτηριστικά που περιγράφουν την κράτηση, όπως ημερομηνίες, τύπο δωματίου, αριθμό ενηλίκων, παιδιά, και ακυρώσεις.

Στο συγκεκριμένο dataset θα βρείτε 119.390 εγγραφές για ένα ξενοδοχείο πόλης και ένα ξενοδοχείο θέρετρου. Κάθε εγγραφή αντιπροσωπεύει μια κράτηση ξενοδοχείου μεταξύ της 1ης Ιουλίου 2015 και της 31ης Αυγούστου 2017, συμπεριλαμβανομένης της κράτησης που έγινε πραγματικά και της κράτησης που ακυρώθηκε. Οι εγγραφές έχουν την ακόλουθη δομή:

hotel , is_cancelled , lead_time , arrival_date_year , arrival_date_month , arrival_date_week_number , arrival_date_day_of_month , stays_in_weekend_nights , stays_in_week_nights , adults , children , babies , meal , country , market_segment , distribution_channel , is_repeated_guest , previous_cancellations , previous_bookings_not_canceled , reserved_room_type , assigned_room_type , booking_changes , deposit_type , agent , company , days_in_waiting_list , customer_type , adr , required_car_parking_spaces , total_of_special_requests , reservation_status , reservation_status_date

Ζητούμενα:

Να κατασκευαστεί πρόγραμμα σε python που θα απαντά στα ερωτήματα της άσκησης που αναφέρονται στη συνέχεια, και θα παρουσιάζει μενού επιλογών για την προβολή τους, σε ένα γραφικό περιβάλλον.

- Ποια είναι τα βασικά στατιστικά στοιχεία για τις κρατήσεις σε κάθε ξενοδοχείο (μέσος όρος διανυκτερεύσεων, ποσοστό ακυρώσεων)?
- Ποια είναι η κατανομή των κρατήσεων ανά μήνα και ανά εποχή;
- Ποια είναι η κατανομή των κρατήσεων ανά τύπο δωματίου;
- Πόσες κρατήσεις αφορούν οικογένειες, ζευγάρια ή μεμονωμένους ταξιδιώτες;
- Ποιες είναι οι τάσεις των κρατήσεων με την πάροδο του χρόνου;
- Υπάρχει εποχικότητα στις κρατήσεις ή στις ακυρώσεις σε κάθε ξενοδοχείο;

Όλα τα ερωτήματα θα πρέπει να αναπαρασταθούν και γραφικά με διαγράμματα. Τα αποτελέσματα θα χρειαστεί να εισαχθούν σε αντίστοιχους πίνακες μιας βάσης MySQL ή SQLite, και να εξαχθούν σε αντίστοιχα αρχεία .csv (τα .csv θα αποτελούν επίσης παραδοτέα της άσκησης).

Προσοχή όλες οι απαραίτητες ενέργειες μετασχηματισμού των δεδομένων θα πρέπει να γίνουν αποκλειστικά σε γλώσσα Python. Η βάση δεδομένων θα χρησιμοποιηθεί **ΜΟΝΟ** για αποθήκευση.

Υπόδειξη: Χρησιμοποιείτε τις απαραίτητες βιβλιοθήκες της Python όπως pandas, xlrd , matplotlib ή όποιες άλλες κρίνετε εσείς απαραίτητο.

Παραδοτέα

1. **Γραπτή Αναφορά** (σε αρχείο pdf ή word) που θα περιλαμβάνει:
 - Τον κώδικα σε γλώσσα python εμπλουτισμένο με σχόλια
 - Screenshots παραδειγμάτων της εφαρμογής (και του σχήματος της βάσης δεδομένων)
 - Τα ζητούμενα γραφήματα (με τίτλους, υπόμνημα)
 - Σχόλια - Παραδοχές που έγιναν για την ανάπτυξη της εργασίας
2. Συμπιεσμένα σε **ένα αρχείο zip**:
 - Την πιο πάνω γραπτή αναφορά
 - Τον **ΤΕΛΙΚΟ** κώδικα σε python.
 - Export του σχήματος, data της βάσης δεδομένων
 - Τα αρχεία .csv που θα περιέχουν τα αποτελέσματα

Το αρχείο zip πρέπει να έχει όνομα τον αριθμό μητρώου του φοιτητή (π.χ. **3972.zip**), και να ανεβεί (**ΥΠΟΧΡΕΩΤΙΚΑ**) στο e-class. Σε ξεχωριστό αρχείο .txt μέσα στο zip να αναφέρεται το ονοματεπώνυμο, το έτος, ο αριθμός μητρώου και η e-mail διεύθυνση του φοιτητή.

Διευκρινήσεις

1. Η άσκηση είναι **ατομική**
2. Η άσκηση είναι **προαιρετική με bonus έως 1 μονάδα στον τελικό βαθμό, εφόσον ο βαθμός της είναι ≥ 5**
3. Τελική ημερομηνία παράδοσης είναι η ημερομηνία της γραπτής εξέτασης Ιουνίου 2024 **ΜΟΝΟ!** Αναλόγως θα καθοριστεί και η ημερομηνία της προφορικής εξέτασης της εργασίας.
4. Η άσκηση **ΔΕΝ ΘΑ ΚΡΑΤΗΘΕΙ** για το επόμενο ΑΚΑΔΗΜΑΙΚΟ ΕΤΟΣ!
5. Για τυχόν απορίες ή υποδείξεις μπορείτε να απευθύνεστε με e-mail στο mvonitsanos@ceid.upatras.gr ή στο χώρο «Συζητήσεις» στη σελίδα του μαθήματος στο eclass
(<https://eclass.upatras.gr/modules/forum/?course=CEID1091>)