

**Αρχές Γλωσσών Προγραμματισμού & Μεταφραστών**

Τμήμα Μηχανικών Η/Υ &amp; Πληροφορικής

Πανεπιστήμιο Πατρών

Εαρινό Εξάμηνο 2023

Διδάσκοντες: Ιωάννης Γαροφαλάκης, Σπυρίδων Σιούτας, Παναγιώτης Χατζηδούκας

**Προαιρετική Εργαστηριακή Άσκηση Python**

Το οικοσύστημα Python που περιλαμβάνει τη γλώσσα προγραμματισμού καθώς και μια σειρά από εξωτερικά πακέτα και βιβλιοθήκες λογισμικού είναι 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.

Η γλώσσα Python 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.stats.govt.nz/large-datasets/csv-files-for-download/> υπάρχουν δεδομένα για ανάλυση/επεξεργασία σε αρχεία csv που αφορούν τους ακόλουθους τομείς: Business, Census, Economy, Effects of COVID-19 on trade, Environment, Government finance, Health, Industries, Labour market, Population, Society.

Στη συγκεκριμένη εργαστηριακή άσκηση θα ασχοληθούμε με δεδομένα του **Effects of COVID-19 on trade**, και συγκεκριμένα το **“effects-of-covid-19-on-trade-at-15-december-2021-provisional.csv”**<sup>1</sup>, που περιέχει συνολικά 111.438 εγγραφές, με την ακόλουθη δομή:

Direction | Year | Date | Weekday | Country | Commodity | Transport\_Mode | Measure | Value | Cumulative

Το συγκεκριμένο αρχείο περιέχει τα ακόλουθα πεδία:

- Direction (*imports, exports, reimports*)
- Year (*2015, ..., 2021*)
- Date (*01/01/2015, 01/02/2015, ..., 15/12/2021*),
- Weekday (*Monday, ..., Sunday*)
- Country (*All, China, European Union, Asia, Australia, USA, ...*)
- Commodity (*“All”, “Milk powder, butter, and cheese”, “Fish, crustaceans, and molluscs”, “Non-food manufactured goods”, “Electrical machinery and equip”, ...*)
- Transport\_Mode (*All, sea, air, ...*)

<sup>1</sup> Stats NZ and licensed by Stats NZ for reuse under the Creative Commons Attribution 4.0 International licence.

- Measure (\$, Tonnes)
- Value (long integer)
- Cumulative (long integer)

### **Ζητούμενα:**

Αρχικά να υλοποιηθεί python script που θα συγκεντρώνει τα δεδομένα από την παραπάνω ιστοσελίδα. Στη συνέχεια να κατασκευαστεί πρόγραμμα σε python που θα επεξεργάζεται μόνο τα απαραίτητα δεδομένα σχετικά με τα ζητούμενα γραφήματα της άσκησης και θα παρουσιάζει μενού επιλογών για τη δημιουργία τους, σε ένα γραφικό περιβάλλον.

Συγκεκριμένα θα χρειαστεί να εξαχθούν τα γραφήματα για τις παρακάτω περιπτώσεις:

- Συνολική παρουσίαση του τζίρου (στήλη value) ανά μήνα (στις αντίστοιχες μονάδες μέτρησης)
- Συνολική παρουσίαση του τζίρου (στήλη value) για κάθε χώρα (στις αντίστοιχες μονάδες μέτρησης)
- Συνολική παρουσίαση του τζίρου (στήλη value) για κάθε μέσο μεταφοράς (στις αντίστοιχες μονάδες μέτρησης)
- Συνολική παρουσίαση του τζίρου (στήλη value) για κάθε μέρα της εβδομάδας (στις αντίστοιχες μονάδες μέτρησης)
- Συνολική παρουσίαση του τζίρου (στήλη value) για κάθε κατηγορία εμπορεύματος (στις αντίστοιχες μονάδες μέτρησης)
- Παρουσίαση των 5 μηνών με το μεγαλύτερο τζίρο, ανεξαρτήτως μέσου μεταφοράς και είδους ανακυκλώσιμων ειδών
- Παρουσίαση των 5 κατηγοριών εμπορευμάτων με το μεγαλύτερο τζίρο, για κάθε χώρα
- Παρουσίαση της ημέρας με το μεγαλύτερο τζίρο, για κάθε κατηγορία εμπορεύματος

Στο τέλος τα συγκεκριμένα δεδομένα να φορτωθούν σε αντίστοιχους πίνακες μιας βάσης MySQL (ή όποιας βάσης επιθυμείτε), 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 τελικά δεδομένα

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  $\geq 5$**
3. Τελική ημερομηνία παράδοσης είναι η ημερομηνία της γραπτής εξέτασης Ιουνίου 2023 **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](mailto:mvonitsanos@ceid.upatras.gr) ή στο χώρο «Συζητήσεις» στη σελίδα του μαθήματος στο eclass  
(<https://eclass.upatras.gr/modules/forum/?course=CEID1091>)