ΠΡΟΑΙΡΕΤΙΚΗ ΕΡΓΑΣΙΑ PYTHON

ΡΟΥΧΩΤΑ ΕΙΡΗΝΗ

ΑΜ: 1059654

Download the excel files

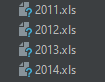
Στο .zip αρχείο που σας έχω αποστείλει θα βρείτε ένα αρχείο **download\_files** το οποίο κατεβάζει αυτόματα τα αρχεία που θέλουμε να επεξεργαστούμε για την εργασίας μας. Ο πιο εύκολος τρόπος που βρήκα είναι μέσω της βιβλιοθήκης wget, στην οποία δηλώνουμε το ulr αρχείο για λήψη και την τοποθεσία αποθήκευσής του. Αρχικά, είχα σκεφτεί μήπως γινόταν με τη χρήση regular expression, αφού όλα τα αρχεία έχουν κοινό το

<a href="[/el/statistics//publication/STO04/](https://www.statistics.gr/el/statistics/-/publication/STO04/2016-Q2)> <divclass="publication selectedPublication"> o Τρίμηνο 201

Στη συγκεκριμένη περίπτωση που τα αρχεία είναι λίγα το υλοποίησα με τη wget. Αν όμως τα αρχεία ήταν πάρα πολλά, σίγουρα θα σκεφτόμασταν την άλλη τεχνική.

Επίσης, για να μην κατεβάζουμε για κάθε χρόνο τέσσερα αρχεία, παρατηρήσαμε ότι στο 4ο τετράμηνο του κάθε χρόνου έχουμε πληροφορίες για όλους τους μήνες.

Mε την εκτέλεση του script, στο τρέχοντα φάκελό μας έχουμε τα παρακάτω αρχεία:

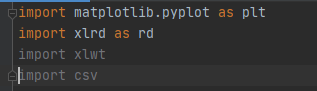


Ζητούμενο 1

Μας ζητήθηκε να φτιάξουμε γράφημα με τις συνολικές αφίξεις τουριστών στην Ελλάδα για τη τετραετία 2011-2014. Αρχικά, πρέπει να αποφασίσουμε ποια στοιχεία θα επιλέξουμε. Ουσιαστικά, για την κάθε χρονολογία, πηγαίναμε στο **αντίστοιχο excel** αρχείο και πιο συγκεκριμένα στο **τελευταίο φύλλο(Δεκέμβριος)** και εξάγουμε τη **τελευταία γραμμή TOTAL ARRIVALS** **και 6η στήλη**. Τα δεδομένα αυτά, αποθηκεύονται σε μία ***δισδιάστατη λίστα****.*



Να πάρουμε ένα παράδειγμα για το έτος 2011

1. Φορτώνουμε όλες τις χρήσιμες βιβλιοθήκες 
2. Διαβάζουμε το αρχείο *2011.xls*



1. Διαβάζουμε το τελευταίο φύλλο του αρχείου



1. Εξάγουμε τη τελευταία γραμμή και 6η στήλη του φύλλου αυτού



Το -2 είναι γιατί ως τελευταία γραμμή θεωρεί ***\* Source: Border Survey of the Bank of Greece.***

1. Δημιουργούμε τον άξονα χ



1. Δημιουργούμε το γράφημα

plt.bar(xAxis,[value2011,value2012,value2013,value2014])

1. Αποθηκεύουμε το γράφημα ως εικόνα *zhtoumeno1.png* στο τρέχοντα φάκελο μάς.



ΓΡΑΦΗΜΑ ΓΙΑ ΖΗΤΟΥΜΕΝΟ 1

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Εικόνα που περιέχει στιγμιότυπο οθόνης, σχεδίαση  Περιγραφή που δημιουργήθηκε αυτόματα |  |  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |  |  |

Ζητούμενο 2

Μας ζητήθηκε να φτιάξουμε γράφημα με τις top 10 χώρες με το μεγαλύτερο μερίδιο στις αφίξεις τουριστών στην Ελλάδα για την τετραετία 2011-2014. Η ιδέα είναι ότι διαβάζουμε πάλι το τελευταίο φύλλου στου excel. Στη συνέχεια πηγαίνουμε στο συγκεντρωτικό πίνακα για όλη τη χρονιά. Διαβάζουμε την 6η στήλη για όλες τις χώρες και για όλες τις χρονιές. Στη συνέχεια για την κάθε χώρα αθροίζουμε το σύνολο των αφίξεων. Αυτά τα δεδομένα τα έχουμε αποθηκευμένα στη λίστα **total\_arrival\_year.** Μέσω μιας δομής επανάληψης βρίσκουμε τις μεγαλύτερες 10 αφίξεις. Για να βρούμε σε ποιες χώρες αντιστοιχούν αυτές παίρνουμε το ***index*** αυτής της λίστας.

Επειδή η μορφοποίηση ήταν τέτοια χωρίσαμε τις «χρήσιμες» γραμμές σε διαστήματα

***78-111 //Europe***

***113-121 //Asia***

***123-125 //Africa***

***127-132 //America***

***134-135 //Oceania***

Πιο αναλυτικά τα βήματα:

1. Φορτώνουμε όλες τις χρήσιμες βιβλιοθήκες



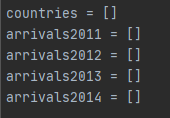
1. Διαβάζουμε το αρχείο *2011.xls*



1. Διαβάζουμε το τελευταίο φύλλο του αρχείου



1. Οι λίστες που θα γεμίσουμε



1. Για παράδειγμα για το 2011 αναλυτικά οι αφίξεις γίνονται μέσω της loop στα διαστήματα (***78-111//Europe 113-121//Asia 123-125 //Africa 127-132 //America 134-135 //Oceania)***

Τα δεδομένα της 6ης στήλης διαβάζονται και γίνονται append στη λίστα **arrivals2011**. Oι χώρες αποθηκεύονται στη λίστα **countries.** Τις χώρες τις πήραμε από το 2014 excel. Θα μπορούσε να γίνει από οποιοδήποτε, αφού η μορφοποίηση είναι ίδια.



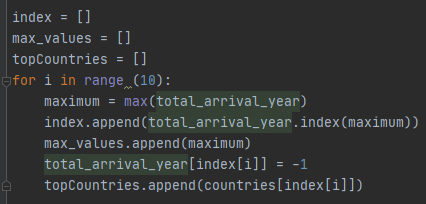
 **ΠΑΡΑΤΗΡΗΣΗ: Η Κροατία, εντάχθηκε στην Ευρωπαϊκή Ένωση το 2013, με συνέπεια να μην υπολογίζεται στα έτη 2013 και 2014 στις υπόλοιπες Ευρωπαϊκές Χώρες. To πρόβλημα αυτό λύθηκε με την εισαγωγή της γραμμής Κροατία στα έτη 2011-2012 με τιμή μηδέν.**



1. Για κάθε χώρα, αθροίζουμε τις αφίξεις της για όλα τα έτη:



1. Βρίσκουμε τις δέκα μεγαλύτερες αφίξεις από τον πίνακα **total\_arrival\_year** και μέσω του index βρίσκουμε τις χώρες που αντιστοιχεί:



MAX\_VALUES

1. Έτσι οι τελικές τιμές είναι:



INDEXES π.χ. 4 αντιστοιχεί στη Γερμανία (η αρίθμηση ξεκινά από το μηδέν)

ΓΡΑΦΗΜΑ ΓΙΑ ΖΗΤΟΥΜΕΝΟ 2

Εικόνα που περιέχει στιγμιότυπο οθόνης

Περιγραφή που δημιουργήθηκε αυτόματα

Ζητούμενο 3

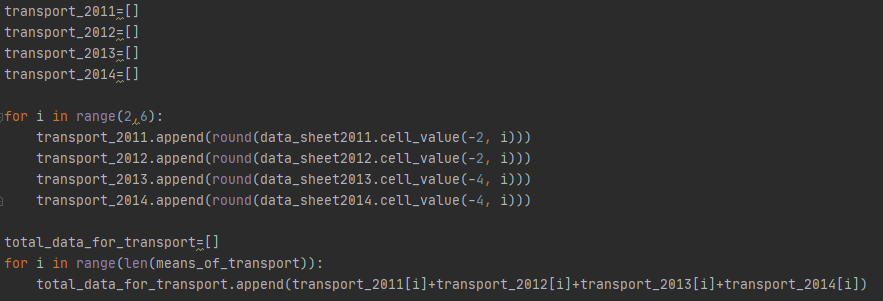
Μας ζητήθηκε να φτιάξουμε γράφημα με αφίξεις των τουριστών στην Ελλάδα ανά μέσο μεταφοράς για τη τετραετία 2011-2014. Ακολουθούμε την ίδια λογική όπως για το πρώτο ζητούμενο, δηλαδή κρατάμε την ίδια γραμμή, απλά τώρα εξάγουμε δεδομένα από τις στήλες **AIR RAILWAY SEA και ROAD.**

*Παραθέτω τα βήματα που είναι διαφορετικά σε σχέση με την υλοποίηση του 1ου ζητούμενου.*

1. Δημιουργούμε άξονα χ.

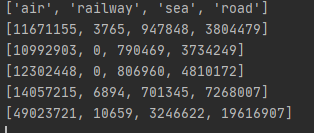


1. Για κάθε χρονιά, παίρνουμε τα συγκεντρωτικά στοιχεία για τα μέσα μεταφοράς. Στη συνέχεια αυτά τα στοιχεία τα αθροίζουμε ένα προς ένα και αποθηκεύονται στη λίστα ***total\_data\_for\_transport***



1. Τα δεδομένα που προκύπτουν είναι:

2011



**ΤΕΛΙΚΟ ΑΠΟΤΕΛΕΣΜΑ**

2014

2013

2012

1. Από το τελικό αποτέλεσμα, είναι φανερό ότι η τάξη μεγέθους για το railway είναι κατά πάρα πολύ μικρότερη από τη τάξη μεγέθους για τα άλλα μεταφορικά μέσα. Προκειμένου να φαίνονται με ευκρίνεια όλα αυτά τα αποτελέσματα και να μην «κρύβονται» τα δεδομένα για το railway, χρησιμοποιήσαμε λογαριθμική κλίμακα.



ΓΡΑΦΗΜΑ ΓΙΑ ΖΗΤΟΥΜΕΝΟ 3

Εικόνα που περιέχει στιγμιότυπο οθόνης

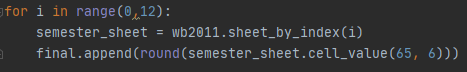
Περιγραφή που δημιουργήθηκε αυτόματα

Ζητούμενο 4

Μας ζητήθηκε να φτιάξουμε γράφημα με αφίξεις των τουριστών στην Ελλάδα ανά τρίμηνο για τη τετραετία 2011-2014. Επειδή δεν υπάρχει η πληροφορία αυτή συγκεντρωτικά, για κάθε αρχείο excel διαβάζουμε όλα τα φύλλα και **συγκεκριμένα τη γραμμή 66 και στήλη 7.** Μη ξεχνάμε ότι η αρίθμηση τη python ξεκινάει από το μηδέν. Τα δεδομένα αυτά αποθηκεύονται στο πίνακα final. Τα δεδομένα μας είναι σε πλήθος 12(μήνες)Χ4(χρόνια)=48 δεδομένα. Μετά για να πάρουμε την πληροφορία για τα τρίμηνα αθροίζουμε ανά 3 δεδομένα1.

Αναλυτικότερα,

1. Μέσω μιας for εξάγουμε τα δεδομένα για όλους τους μήνες του χρόνου (π.χ. για το 2011)



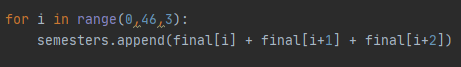
1. Έχοντας κάνει την ίδια διαδικασία και για τους υπόλοιπους χρόνους η λίστα final είναι:







1. Στη συνέχεια, αθροίζουμε ανά τρία για να βρούμε τα συνολικά δεδομένα για τα τρίμηνα στη λίστα **semesters**



1. Τα τελικά δεδομένα ανά τρίμηνο είναι:



ΓΡΑΦΗΜΑ ΓΙΑ ΖΗΤΟΥΜΕΝΟ 4

Εικόνα που περιέχει στιγμιότυπο οθόνης

Περιγραφή που δημιουργήθηκε αυτόματα

PYTHON TO SQLite

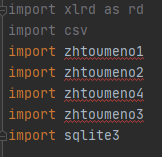
Όπως έχει διατυπωθεί και στην εκφώνηση της εργασίας, τα τελικά αποτελέσματα αποθηκεύονται σε πίνακες SQLite. Η βάση δεδομένων θα χρησιμοποιηθεί μόνο ως χώρος αποθήκευσης.

Τα βήματα που ακολουθήσαμε είναι:

1. Δημιουργούμε μία βάση, η οποία αποθηκεύεται στο φάκελο της εργασίας μας
2. Συνδέουμε τη βάση με το πρόγραμμά μας μέσω ενός connector



1. Θα χρειαστούμε τα δεδομένα από τα προηγούμενα ερωτήματα, για αυτό κάνουμε import τα προγράμματα και όποιες ακόμα βιβλιοθήκες χρειαζόμαστε

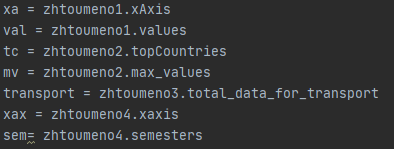


1. Α) Από το ερώτημα 1 εξάγουμε το **xAxis** και **values**

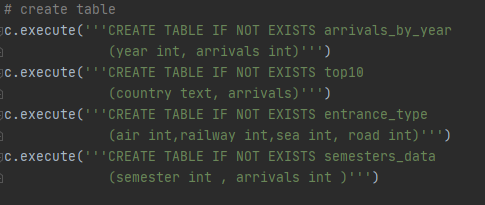
Β) Από το ερώτημα 2 εξάγουμε το **max\_values** και **topCountries**

Γ) Από το ερώτημα 3 εξάγουμε το **total\_data\_for\_transport**

Δ) Από το ερώτημα 4 εξάγουμε το **xAxis** και **semesters**

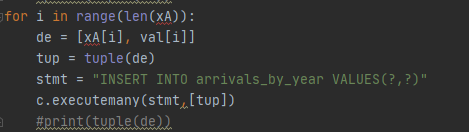


1. Δημιουργούμε τους πίνακες της βάσης. Ουσιαστικά η πληροφορία των γραφημάτων «μεταφέρεται» σε πίνακες βάσεων δεδομένων.

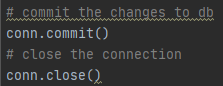


1. Για να κάνω **insert** για παράδειγμα στον πίνακα **arrivals\_by\_year**

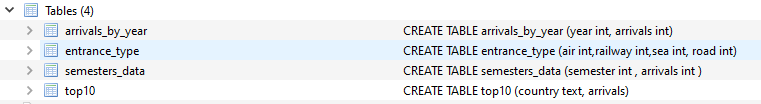
Θεωρούμε ότι το κάθε insert είναι ένα tuple. Παρόμοια και για τους άλλους πίνακες.



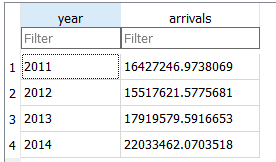
1. Επιβεβαίωση αλλαγών και έξοδος από τη βάση



1. Τα αποτελέσματα στη βάση



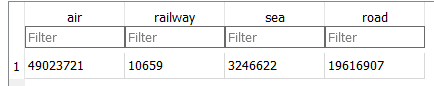
1. Για τον πίνακα **arrivals\_by\_year**



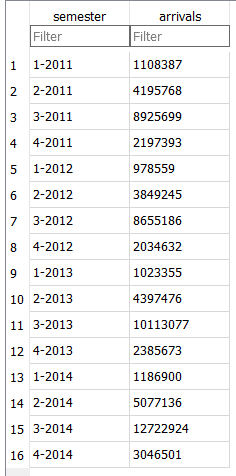
1. Για τον πίνακα **top10**



1. Για τον πίνακα **entrance\_type**



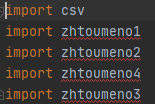
1. Για τον πίνακα **semesters\_data**



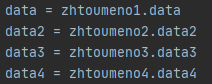
Total Csv

Σε αυτό το σημείο, θέλουμε η πληροφορία των γραφημάτων να «μεταφερθεί» σε ένα αρχείο csv. Όπως και πριν, θα χρειαστούμε δεδομένα από τα προηγούμενα ερωτήματα.

1. Κάνουμε import τις βιβλιοθήκες που χρειαζόμαστε



1. Επιλέγουμε τα δεδομένα που θέλουμε



Όπου: 



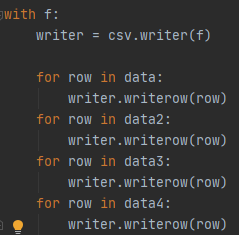




1. Φτιάχνουμε ένα αρχείο csv με δυνατότητα write



1. Κάνουμε εγγραφή των δεδομένων στο αρχείο μέσω του ***csv.writer(f)*** και ***writer.writerow(row)***



1. Το τελικό αρχείο csv περιέχει

