

TD4 : Restaurants new-yorkais

Ce sujet traite des fichiers JavaScript Object Notation (JSON). Ce format permet de stocker des données structurées, par exemple avec une organisation hiérarchique.

Préambule

1. Dans votre dossier PythonOpenData, créez un sous-répertoire TD4 dans lequel vous réaliserez l'ensemble des travaux.
2. Lancer l'éditeur Visual Studio Code depuis Anaconda Navigator.
3. Dans Visual Studio Code, ouvrir le dossier PythonOpenData/TD4.

Dans ce TD, on se tourne vers un jeu de données recensant des notes attribuées à des restaurants de la ville de New York.

Exercice 1 : Chargement des données

1. Sur Cursus, télécharger le jeu de données `NYfood.json` et l'enregistrer dans le dossier PythonOpenData/TD4.
2. Visualiser avec Visual Studio Code le contenu du fichier JSON.
3. Quels sont les attributs (clés de dictionnaires) de ce jeu de données ? Que contient l'attribut `grades` (qui signifie `notes` en français) ?

Exercice 2 : Extraction d'informations élémentaires

Dans la suite de ce sujet, on nomme « restaurant » chaque dictionnaire issu du fichier JSON. Les questions suivantes nécessitent de manipuler les données depuis Python et le code produit devra être inclus dans un nouveau fichier `td4.py`.

1. Charger le contenu du fichier `NYfood.json` dans une liste de dictionnaires, chaque dictionnaire représentant un restaurant.
2. Combien y a-t-il de restaurants dans ce fichier ?
3. Combien y a-t-il de restaurants situés à Manhattan listés dans ce fichier ?
4. Écrire une fonction qui prend en entrée la liste des restaurants et retourne une liste sans doublon des quartiers dans lesquels se situent les restaurants listés dans ce fichier.
5. Écrire une fonction qui prend en entrée la liste des restaurants et affiche, pour chaque restaurant du quartier “Manhattan”, un récapitulatif de la forme `nom_du_restaurant: adresse où l'adresse est composée de building, street, zipcode New York, USA`
6. Quel est le nombre total de notes attribuées aux restaurants du fichier ?
7. Quelles sont les valeurs possibles de notes attribuées aux restaurants par les évaluateurs ?

Exercice 3 : Travail spécifique sur les dates

1. Remarquer que les dates associées aux notes n'ont pas été reconnues comme telles
2. Écrire une fonction qui prend en entrée la liste des restaurants et retourne une version transformée dans laquelle les dates sont codées au format `datetime.datetime`.
3. Calculer, pour chaque mois de l'année 2014, le nombre total de notes attribuées (on rappelle que les variables de type `datetime` ont des attributs `year` et `month` accessibles via `d.year` et `d.month`).

Exercice 4 : Export au format CSV

Dans cette partie, il va s'agir d'enregistrer un export du contenu du fichier `NYfood.json` au format CSV, en s'assurant donc que l'on n'ait plus d'attributs imbriqués.

1. Écrire une fonction qui prend en entrée le jeu de données précédent et retourne une nouvelle version de ce jeu de données dans lequel l'attribut `grades` a été supprimé et remplacé par un attribut `n_grades` qui indique le nombre de notes reçues.
2. Écrire une fonction qui prend en entrée le jeu de données précédent et retourne une nouvelle version de ce jeu de données dans lequel l'attribut `address` est supprimé et remplacé par deux nouveaux attributs `latitude` et `longitude` (l'attribut `coordinates` est une liste contenant la longitude et la latitude, stockées dans cet ordre).
3. En utilisant les deux fonctions précédentes, exporter cette nouvelle version du jeu de données dans un fichier nommé `NYfood.csv`.