

Analyse de Données en python

d'après une activité de Marie Durand

Dans cette activité, nous allons utiliser le langage Python pour effectuer des traitements sur des données.

Répondez aux questions directement sur cette feuille.

À FAIRE 1

Téléchargez le fichier <https://nodfs.xyz/analyse.7z> et extrayez-le (dans l'explorateur de fichier, clic-droit dessus, 7-zip, Extraire Ici).

1 Premiers pas

Le fichier «au_marche.csv» contient les mêmes données que celles que vous avez utilisées dans la première activité à distance de la semaine dernière.

QUESTION 1

Ouvrez le fichier avec le Bloc-Note.
Par quoi sont séparées les valeurs sur une même ligne ?
Que contient la première ligne ?

À FAIRE 2

Utilisez un logiciel permettant d'éditer du code Python, par exemple Edupython ou Pyzo.
Chargez et exécutez le fichier fruits_legumes_fleurs.py

Vous devez obtenir les mêmes éléments que dans l'image suivante (avec les prix en plus) affichés au format CSV (pas dans un tableau comme ici) :

Nom	Type	Couleur
Pomme	fruit	rouge
Cerise	fruit	rouge
Radis	légume	rouge
Rose	fleur	rouge
Fraise	fruit	rouge

Voici le code utilisé dans le fichier fruits_legumes_fleurs.py :

```
1 from manipulation_csv import *
2 objets = lire_fichier("au_marche.csv", ',')
3 #Filtre pour obtenir les objets de couleur rouge dans la variable objets_rouges
4 objets_rouges = filtrer(objets, 'Couleur', 'rouge')
5 print("Voici les objets de Couleur rouge :")
6 afficher_collection(objets_rouges)
```

Explication de chaque ligne :

ligne 1 À la première ligne on importe toutes les fonctions contenues dans le fichier manipulation_csv.py.

- **ligne 2** On utilise la fonction lire_fichier qui permet de lire le contenu d'un fichier, ici au_marche.csv dont le séparateur est ','. Le contenu du fichier, donc la collection de tous les objets décrits dans le fichier, est affecté à la variable objets.
- **ligne 3** La ligne 3 commence par le symbole # et s'affiche dans une couleur différente : il s'agit d'un commentaire.
- **ligne 4** On utilise la fonction filtrer qui permet d'effectuer un filtre sur une collection d'objets pour ne conserver que ceux dont le descripteur a la valeur demandée. La collection objets n'est pas modifiée par cet appel, donc on peut toujours l'utiliser ensuite.
- **ligne 5** On utilise la fonction print() pour afficher un texte sur la console.
- **ligne 6** La fonction afficher_collection permet d'afficher le contenu de la collection d'objets passée en paramètre.

À FAIRE 3

Le programme affiche seulement les objets de couleur rouge.
Utilisez de nouveau la fonction afficher_collection, mais pour afficher tous les objets.

INFO 1

Une **fonction** est un bloc d'instructions auquel on a donné un nom. Ce bloc d'instructions pourra être exécuté en **appelant** la fonction, c'est-à-dire en utilisant son nom. Les fonctions peuvent avoir des paramètres. Un paramètre d'une fonction est une valeur transmise à la fonction lors de son appel. Les paramètres sont transmis dans un ordre fixé, à l'intérieur des parenthèses, après le nom de fonction.

Par exemple, la fonction afficher_collection prend pour paramètre une collection d'objets, objets_rouges et affiche son contenu. On peut l'utiliser aussi bien avec la variable objets qu'avec la variable objets_rouges.

À FAIRE 4

Créez une nouvelle variable objets_verts et utilisez la fonction filtrer() de manière à ce que objets_verts contienne les objets de couleur verte. Affichez ces objets.

QUESTION 2

Quel filtre faut-il appliquer pour obtenir les objets de la liste suivante ?

Nom	Type	Couleur	Prix
Pomme	fruit	rouge	3€
Concombre	légume	vert	3€
Kiwi	fruit	vert	3€
Salade	légume	vert	3€
Orange	fruit	orange	3€
Rose	fleur	rouge	3€

À FAIRE 5

Ajoutez à la suite de votre programme les lignes nécessaires pour sélectionner les objets correspondant à l'image ci-dessus et affichez-les pour vérifier que vous obtenez les objets attendus.

2 Application

L'INSEE est l'institut chargé de la production, de l'analyse et de la publication des statistiques officielles en France.

Nombre de ces données sont accessibles librement sur son site, par exemple à l'adresse <https://www.insee.fr/fr/statistiques/4994589> on peut télécharger un fichier CSV de 2.3 millions de ligne, chacune listant des informations sur un-e salarié-e, l'ensemble représentant 1/12e des salarié-e-s en France en 2017.

Pour faciliter un peu le traitement, je vous ai préparé le fichier `salaires_2017_20k.csv` qui contient seulement 20000 lignes prises au hasard, mais vous pouvez très bien télécharger ensuite le fichier complet et faire les mêmes opérations dessus (il pèse 216Mo et cela prendra moins d'une minute pour exécuter le programme python dessus).

QUESTION 3

Ouvrez dans le Bloc-Note le fichier `salaires_2017_20k.csv`. Les noms des champs sur la première ligne ainsi que les valeurs de ces champs sur les lignes suivantes sont un peu mystérieux, et sont détaillés dans un autre fichier CSV : `Varmod_SALAAAN_2017.csv`.

En regardant les deux fichiers, trouvez dans quelle tranche est le salaire annuel de la personne de la première ligne.

À FAIRE 6

Ouvrez dans Pyzo ou Edupython le fichier `analyse_insee.py` et exécutez le.

À FAIRE 7

La fonction `len()` en python renvoie le nombre d'éléments dans l'objet qu'on lui passe en paramètre (donc le nombre de lignes si on lui envoie une collection).

Rajoutez une variable «hommes» et affichez le nombre d'hommes dans la liste des salariés.

À FAIRE 8

Le champ qui correspond au salaire s'appelle «TRNNETO» et prend des valeurs entre 0 et 23. En refaisant un filtre sur la variable `femmes`, affichez le nombre de femmes du fichier qui ont un salaire dans la tranche la plus élevée (23).

Trouvez de même le nombre d'hommes qui ont un salaire dans cette tranche.

À FAIRE 9

Décommentez (enlevez le «#» au début de chaque ligne) les lignes de la fin du fichier, et exécutez le pour afficher un graphique qui représente combien il y a de femmes et d'hommes (avec des barres superposées en violet et gris) pour chaque tranche de salaire (de la plus basse 0, à la plus haute 23).

QUESTION 4

Que penser de l'égalité salariale entre les femmes et les hommes en France ?