

DL #5: Exercice de nettoyage de données

Devoir libre pour un cours dispensé à l'université de Rennes 2

Romain Tavenard

Le module `csv` fournit, en Python, un ensemble de fonctions permettant de lire et d'écrire efficacement des fichiers de données au format CSV. L'objet de ce devoir libre est de créer un programme de nettoyage de fichiers de données se basant sur ce module `csv`. Vous allez implémenter une fonction qui prend en entrée un nom de fichier (contenant des données au format CSV), le délimiteur utilisé dans ce fichier (virgule, point virgule, tabulation, etc.), une chaîne de caractères désignant une donnée manquante dans le fichier et, de manière optionnelle, un nom de fichier de sortie dans lequel écrire le contenu nettoyé (si ce nom n'est pas fourni, le contenu nettoyé sera affiché dans la console).

Cette fonction devra vérifier, pour chaque donnée dans le fichier, si celle-ci est numérique. Si c'est le cas, la donnée sera conservée, sinon, elle sera remplacée par la chaîne de caractères correspondant aux données manquantes. Par exemple, si le fichier d'entrée contient le texte suivant :

```
X0;X1;X2
12;15;2.5
55;MANQUANT;8
3;2.8;
```

et si le délimiteur choisi est ";" et la chaîne indiquant les données manquantes est "NA", le programme devra écrire :

```
X0;X1;X2
12;15;2.5
55;NA;8
3;2.8;NA
```