

TP : Manipulation de données avec pandas

Prof : Mr SY

Master 2 : IAGE

School : ISI

Exercice 1 : Importation et exportation

1. Télécharger à la main le fichier csv à l'adresse suivante : <http://egallic.fr/Enseignement/Python/Exercices/donnees/notes.csv> et le placer dans le répertoire courant. Importer son contenu dans Python.
2. Importer à nouveau les données dans Python, mais en fournissant cette fois le l'url directement à la fonction d'importation.
3. À présent, importer le contenu du fichier disponible à l'adresse http://egallic.fr/Enseignement/Python/Exercices/donnees/notes_decim.csv. Le séparateur de champs est un point virgule et le séparateur décimal est une virgule.
4. Importer le contenu du fichier http://egallic.fr/Enseignement/Python/Exercices/donnees/notes_h.csv. Le nom des colonnes n'est pas présent.
5. Importer le contenu du fichier http://egallic.fr/Enseignement/Python/Exercices/donnees/notes_h_s.csv. La première ligne n'est pas à importer.
6. Importer le contenu de la première feuille du fichier Excel <http://egallic.fr/Enseignement/Python/Exercices/donnees/notes.xlsx>.
7. Importer le contenu de la seconde feuille (notes_h_s) du fichier Excel disponible ici : <http://egallic.fr/Enseignement/Python/Exercices/donnees/notes.xlsx>. La première ligne est un commentaire à ne pas considérer durant l'importation.
8. Exporter le contenu de l'objet contenant les notes de la question précédente au format csv (virgule en séparateur de champs, point en séparateur décimal, ne pas conserver le numéro des lignes).

Exercice 2 : Manipulation de tableaux de données

1. À l'aide de la fonction `read_excel()` de la librairie `pandas`, importer le contenu de la feuille intitulée `notes_2012` du fichier Excel disponible à l'adresse suivante : http://egallic.fr/Enseignement/Python/Exercices/donnees/notes_etudiants.xlsx et le stocker dans une variable que l'on nommera `notes_2012`.
2. Afficher les 6 premières lignes du jeu de données, puis les dimensions du tableau.
3. Conserver uniquement la colonne `note_stat` du tableau de données `notes_2012` dans un objet que l'on appellera `tmp`.
4. Conserver uniquement les colonnes `num_etudiant`, `note_stat` et `note_macro` dans un objet nommé `tmp`.
5. Remplacer le contenu de `tmp` par les observations de `notes_2012` pour lesquelles l'individu a obtenu une note de stat supérieure (strictement) à 10.

6. Remplacer le contenu de `tmp` par les observations de `notes_2012` pour lesquelles l'individu a obtenu une note de stats comprise dans l'intervalle (10, 15).
7. Regarder s'il y a des doublons dans le tableau de données `notees_2012` ; le cas échéant, les retirer du tableau.
8. Afficher le type des données de la colonne `num_etudiant`, puis afficher le type de toutes les colonnes de `notes_2012`.
9. Ajouter au tableau `notes_2012` les colonnes suivantes :
 1. `note_stat_maj` : la note de stat (`note_stat`) majorée d'un point,
 2. `note_macro_maj` : la note de macro (`note_macro`) majorée de trois points (le faire en deux étapes : d'abord deux points en plus, puis un point).
10. Renommer la colonne `year` en `annee`.
11. Depuis le fichier `notes_etudiants.xlsx` (c.f. question 1), importer le contenu des feuilles `notes_2013`, `notes_2014` et `prenoms` et le stocker dans les objets `notes_2013`, `notes_2014` et `prenoms`, respectivement.
12. Empiler le contenu des tableaux de données `notes_2012`, `notes_2013` et `notes_2014` dans un objet que l'on nommera `notes`.
13. Fusionner les tableaux `notes` et `prenoms` à l'aide d'une jointure gauche, de manière à rajouter les informations contenues dans le tableau `prenoms` aux observations de `notes`. La jointure doit se faire par le numéro d'étudiant et l'année, l'objet final viendra remplacer le contenu de `notes`.
14. Trier le tableau `notes` par années croissantes et notes de macro décroissantes.
15. Créer une colonne `apres_2012` qui prend la valeur `True` si l'observation concerne une note attribuée après 2012.
16. En effectuant des regroupements sur le dataframe `notes` calculer :
 1. la moyenne et l'écart-type annuels des notes pour chacune des deux matières,
 2. la moyenne et l'écart-type annuels et par sexe des notes pour chacune des deux matières.