

Activité 4 sur le thème 1 : traitement de données à l'aide de Python

Consignes :

1. Enregistrer immédiatement le fichier word dans le dossier approprié (snt/theme1).
2. Compléter ce document Word . Il faudra soigner la rédaction (**phrases complètes**). A la fin de l'heure , enregistrer votre travail sur clé usb et déposer le document dans l'espace élève de Pronote. (page d'accueil ou cahier de texte – onglet déposer ma copie)

La bibliothèque pandas est utilisée pour la structuration des données . Elle permet notamment d'exploiter des données sous format csv

<code>import pandas</code>	nous importons la bibliothèque pandas afin de pouvoir l'utiliser
<code>table=pandas.read_csv("tableau.csv", delimiter=";")</code>	nous créons une variable "table" qui va contenir les données présentes dans le fichier "tableau.csv"
<code>requete1=table.loc[1,'descripteur1']</code>	requete1 stocke la valeur du descripteur1 correspondant à la ligne 1 de la base de donnée « table »
<code>requete2=table.loc[:,["descripteur2"]]</code>	requete2 stocke les valeurs du descripteur2 correspondant à toute la base de donnée « table »
<code>requete3=table.loc[table["descripteur numerique4"]>... ,["descripteur3"]]</code>	requete3 stocke les valeurs du descripteur3 avec une contrainte portant sur le descripteur4 correspondant à toute la base de donnée « table »
<code>tablefusion=pandas.merge(table1, table2)</code>	« tablefusion » est la base de données correspondant à la fusion de la « table1 » et de la « table2 » lorsque celles-ci ont une colonne en commun.
<code>tablefusion.to_csv('fusion.csv', index=False)</code>	Permet d'enregistrer la nouvelle table de donnée sous le com « fusion.csv »

Exercice 1 : la bibliothèque Panda – traitement de données

On se donne les tables de données ci-dessous :

Table 1

Nom	Prenom	Annee	Profession
BREGNON	Jean	1973	Charcutier
BACHOT	Sophie	1978	Chef d'entreprise
DURAND	Philippe	1950	Retraite
TRICOT	Esteban	1991	Traducteur
LAMERINE	Celia	1998	Architecte
BECOUZE	Ludivine	1967	Enseignant
LEOTAN	Yann	1995	Informaticien

Table 2

Nom	Prenom	Sport	Instrument
BREGNON	Jean	Escalade	Aucun
BACHOT	Sophie	Equitation	Piano
DURAND	Philippe	Petanque	Accordeon
TRICOT	Esteban	Course a pied	Aucun
LAMERINE	Celia	Danse	Trompette
BECOUZE	Ludivine	Basket-Ball	Piano
LEOTAN	Yann	Badminton	Guitare

Les tables ci-dessus sont disponibles au format CSV à l'aide des liens : mathssa.fr/table1.csv et mathssa.fr/table2.csv

1. Récupérer les fichiers table1.csv et table2.csv et les enregistrer dans snt/theme1. Ouvrir un nouveau module d'EDUPYTHON et l'enregistrer sous le nom exercice1 toujours dans le dossier snt/theme1.

2. Ecrire les instructions suivantes dans l'éditeur d'EDUPYTHON. (on pourra faire du copier-coller)

```
import pandas
```

```
table1=pandas.read_csv("table1.csv",delimiter=";")
```

```
table2=pandas.read_csv("table2.csv",delimiter=";")
```

Exécuter le programme puis dans la console faire afficher table1 et table2.

3. Dans l'éditeur, écrire **à la suite** :

```
requete1=table1.loc[ :, ["Nom"]]
```

```
requete2=table2.loc[2, :] Que contient requete1 ?
```

.....

```
Que contient requete2 ? .....
```

4. Dans l'éditeur, écrire **à la suite** une instruction que l'on appellera requete3 permettant d'afficher les professions exercées par les personnes associées à ces bases de données.

.....

4. Compléter l'instruction suivante afin que requete4 contienne le résultat de la requête suivante :

« Sélectionner Nom dans table1 avec Année >1990 »

requete4=table1.loc[table1["....."]>....,["....."]]

Que renvoie la requête ?

5. Compléter l'instruction suivante afin que requete5 renvoie le nom des personnes de la table2 pratiquant le piano.

requete5=.....loc[.....["....."]== ".....",["....."]]

Que renvoie la requête ?

6. Compléter l'instruction suivante afin que requete6 renvoie la profession des personnes pratiquant le piano.

requete6=.....loc[.....["....."]== ".....",["....."]]

Faire une capture d'écran de l'éditeur de programme et de la console et la coller ci-dessous.

7. Compléter les instructions suivantes afin de fusionner les collections table1 et table2

tablefusion=pandas.merge(.....,.....)
tablefusion.to_csv('fusion.csv', index=False)

Ouvrir le fichier fusion.csv et vérifier que les tables ont bien fusionné.

Exercice 2 : Analyse des naissances par département

1. Télécharger et enregistrer le fichier mathssa.fr/prenoms.csv dans le dossier snt/theme1

2. Ouvrir un nouveau module d'EDUPYTHON et l'enregistrer sous le nom exercice2 toujours dans le dossier snt/theme1.

3. Ecrire les instructions suivantes dans l'éditeur d'edupython. (on pourra faire du copier coller)

import pandas
prenoms=pandas.read_csv("prenoms.csv",delimiter=";")

4. Exécuter le programme puis dans la console faire afficher prenom

5. La 1^{ère} ligne affichée correspond aux descripteurs, quels sont-ils ?

.....

6.

```
• 1 import pandas
• 2 prenomns=pandas.read_csv("prenomns.csv",delimiter=";")
• 3 recherche=prenomns.loc[(prenomns["preusuel"]=="GILBERT")& (prenomns["annais"]=="1960")]
• 4 print(recherche)
```

Que permet d'obtenir ce programme ? (ne pas l'écrire)

.....
.....

7.Recopier et exécuter le programme pour vérifier la réponse. On pourra copier-coller ce qui suit :

```
import pandas
prenomns=pandas.read_csv("prenomns.csv",delimiter=";")
recherche=prenomns.loc[(prenomns["preusuel"]=="GILBERT")&(prenomns["annais"]=="1960")]
print(recherche)
```

8.Ecrire un programme permettant de déterminer combien de personnes ayant le même prénom que vous sont nées la même année dans le département de l'Allier.