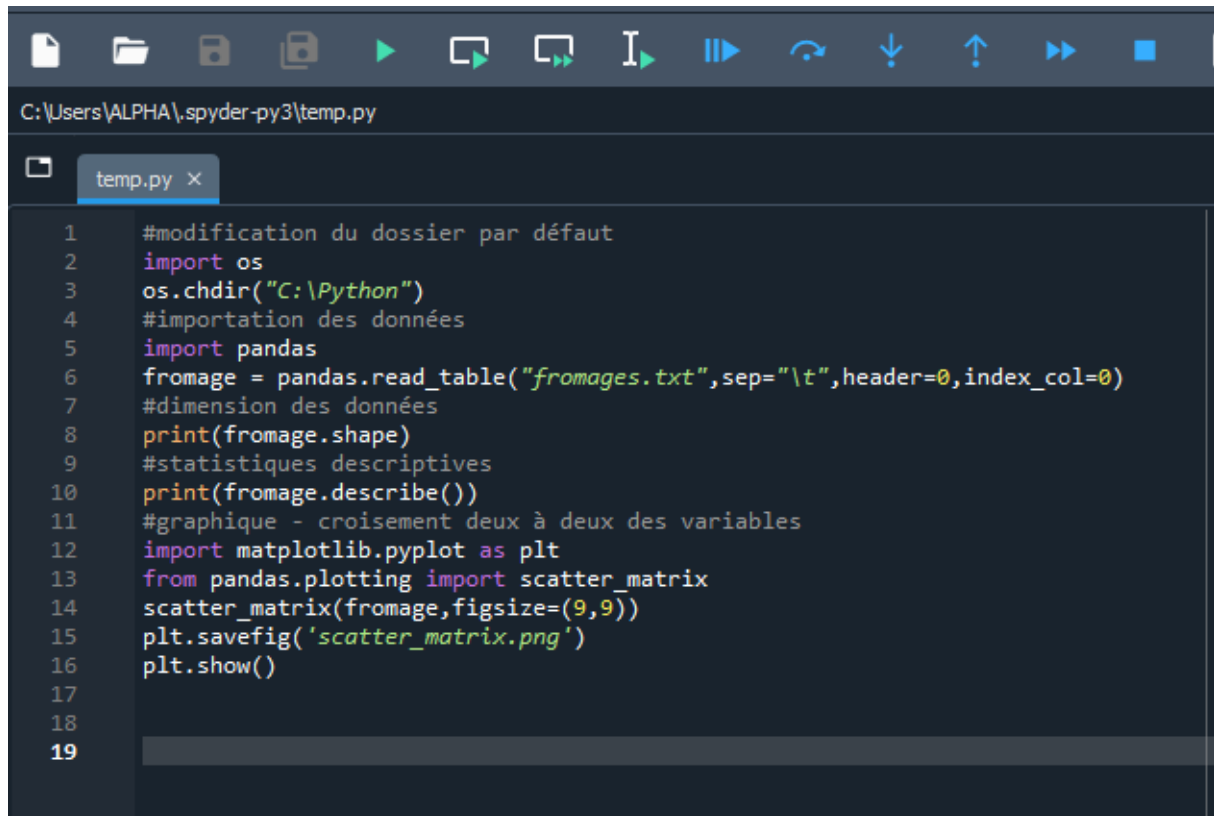


TP1 : Activité sous Spyder de Anaconda



```
1 #modification du dossier par défaut
2 import os
3 os.chdir("C:\\Python")
4 #importation des données
5 import pandas
6 fromage = pandas.read_table("fromages.txt",sep="\t",header=0,index_col=0)
7 #dimension des données
8 print(fromage.shape)
9 #statistiques descriptives
10 print(fromage.describe())
11 #graphique - croisement deux à deux des variables
12 import matplotlib.pyplot as plt
13 from pandas.plotting import scatter_matrix
14 scatter_matrix(fromage,figsize=(9,9))
15 plt.savefig('scatter_matrix.png')
16 plt.show()
17
18
19
```

1- Importation et Modification de modules :

- **os.chdir** : effectuer des opérations liées au système d'exploitation, telles que la modification du répertoire de travail « **C:\\Python** » Cela signifie que tous les fichiers et données utilisés dans la suite du code seront recherchés dans ce répertoire.
- **pandas** : bibliothèque pour la manipulation et l'analyse de données tabulaires.
- **matplotlib.pyplot**: module pour créer des graphiques et des visualisations de données.
- **pandas.plotting.scatter_matrix**: création d'une matrice de dispersion des données.

2- Lecture des données :

- Le code utilise la bibliothèque pandas pour lire un fichier nommé **"fromages.txt"**.
- Les données dans ce fichier sont **tabulées** (séparées par des tabulations) et ont une ligne d'en-tête (**header**).
- Les données sont chargées dans un **DataFrame** fromage, qui est essentiellement une structure de données tabulaire.

3- Affichage des informations sur les données :

- `print(fromage.shape)`: Cette ligne affiche la forme du DataFrame fromage, c'est-à-dire le nombre de lignes et de colonnes.

```
In [1]: runfile('C:/Users/ALPHA/.spyder-py3/temp.py', wdir='C:/Users/ALPHA/.spyder-py3')
(29, 9)
```

- `print(fromage.describe())`: Cette ligne affiche un résumé statistique des données du DataFrame, y compris des statistiques telles que la moyenne, l'écart-type, les valeurs minimales et maximales, etc.

```
In [3]: runfile('C:/Users/ALPHA/.spyder-py3/temp.py', wdir='C:/Users/ALPHA/.spyder-py3')
(29, 9)
```

	calories	sodium	calcium	...	protéines	cholesterol	magnesium
count	29.000000	29.000000	29.000000	...	29.000000	29.000000	29.000000
mean	300.034483	210.086207	185.734483	...	20.168966	74.586207	26.965517
std	91.914356	108.678923	72.528882	...	6.959788	28.245755	11.318388
min	70.000000	22.000000	72.600000	...	4.100000	10.000000	10.000000
25%	292.000000	140.000000	132.900000	...	17.800000	70.000000	20.000000
50%	321.000000	223.000000	202.300000	...	21.000000	80.000000	26.000000
75%	355.000000	276.000000	220.500000	...	23.400000	90.000000	30.000000
max	406.000000	432.000000	334.600000	...	35.700000	120.000000	51.000000

[8 rows x 9 columns]

4- Création d'une matrice de dispersion :

- Le code utilise la fonction `scatter_matrix` du module `pandas.plotting` pour créer une matrice de dispersion des données dans le DataFrame fromage.
- Cette matrice de dispersion affiche des graphiques de dispersion croisée pour chaque paire de variables dans les données.

5- Sauvegarde et affichage de la matrice de dispersion :

- `plt.savefig('scatter_matrix.png')`: Cette ligne sauvegarde la matrice de dispersion en tant qu'image avec le nom de fichier "scatter_matrix.png".
- `plt.show()`: Cette ligne affiche la matrice de dispersion à l'écran.



Ce code lit des données tabulaires à partir d'un fichier « fromages.txt », effectue une analyse statistique et crée une matrice de dispersion pour explorer la relation entre différentes variables dans les données, puis sauvegarde cette matrice de dispersion sous forme d'image.

