

DOCUMENTATION DE LA BIBLIOTHÈQUE PANDAS

```
import pandas as pd # pandas : gestion de données
```

Voici une liste de commandes utiles pour ce TP. Le tableau des données importées sera ici noté `df` (dataframe) et `Colonne` est le *nom* d'une colonne de ce tableau.

Attention : adapter le `df` en fonction du nom donné à votre tableau au moment de l'import.

<code>df = pd.read_csv(fichier)</code>	import d'un fichier csv.
<code>df = pd.read_excel(fichier)</code>	import d'un fichier excel
<code>df</code>	aperçu du tableau
<code>df.head()</code> , <code>df.head(n)</code>	5 premières lignes, <i>n</i> premières lignes
<code>df.tail()</code> , <code>df.tail(n)</code>	5 dernières lignes, <i>n</i> dernières lignes
<code>df.shape</code>	taille du tableau (lignes, colonnes)
<code>df.info()</code>	affiche les colonnes et leur type
<code>df[Colonne]</code>	données de la colonne <code>Colonne</code> . Attention, si le nom de la colonne est une chaîne de caractères, on écrira son nom entre guillemets.
<code>df[[Colonne1,Colonne2,...]]</code>	sélection des colonnes indiquées.
<code>df[df[Colonne] == 5]</code>	applique un filtre, ici la valeur de <code>Colonne</code> doit être égale à 5.
<code>df[(filtre1) & (filtre2)]</code>	applique plusieurs filtres (booléens), le <code>&</code> signifie 'et'
<code>df.sort_values(Colonne)</code>	trie la <code>Colonne</code> par ordre croissant.
<code>df.describe()</code>	statistiques de base
<code>df.mean()</code>	moyenne
<code>df.std()</code>	écart-type
<code>df.count()</code>	nombre de valeurs
<code>df.median()</code>	médiane
<code>df.max()</code>	maximum
<code>df.min()</code>	minimum
<code>df.sum()</code>	somme des données
<code>df.hist(column = Colonne)</code>	histogrammes de <code>Colonne</code> . Options : <code>range = (a,b)</code> (intervalle $[a, b]$ en abscisse), <code>bins = n</code> (nombre de classes)
<code>df.hist(column = Colonne1, by = Colonne2)</code>	histogrammes de <code>Colonne1</code> pour chaque valeur de <code>Colonne2</code> .