

## Aide-mémoire

### Analyse de données avec Pandas

#### Importation de la bibliothèque Pandas

```
import pandas as ps ou from pandas import *
import matplotlib.pyplot as plt
```

#### Création d'un DataFrame

```
df=DataFrame({'Id_Colonne1':[Val1,...,ValN], ...,
'Id_ColonneM':[Val1,...,ValN]})
```

#### Création d'un DataFrame (importation)

**A partir d'un fichier csv :**

```
df=read_csv("Fichier.csv", sep =" séparateur")
```

**A partir d'un fichier texte :**

```
df=read_table("Fichier.txt", sep =" séparateur")
```

**A partir d'un fichier Excel :**

```
df = read_excel("Fichier.xls", "Nom_Feuille")
```

#### Information de base d'un DataFrame

df.shape	Les dimensions
----------	----------------

df.size	Le nbre d'éléments
---------	--------------------

df.info()	Les informations
-----------	------------------

df.describe()	Un résumé rapide
---------------	------------------

df.columns	Les noms des colonnes
------------	-----------------------

L'affichage (extraire des données)	
print(df)	Le contenu d'un dataframe
df	
df.Nom_colonne	Le contenu d'une colonne
df[ "Nom_colonne" ]	
df[ [ "col1", "col2", ... ] ]	Le contenu de plusieurs colonnes
df.head(n)	Les n premières lignes
df.tail(n)	Les n dernières lignes
<b>Les données des lignes :</b>	
df[ligne_initiale : ligne_finale]	
<b>Les données en utilisant les indices :</b>	
df.iloc[lg_i : lg_f , col_i : col_f]	
<b>Les données des cellules en utilisant des étiquettes :</b>	
df.loc[lg_i : lg_f , col_i : col_f]	
<b>Les données selon une condition :</b>	
df[ df [ "nom colonne" ] op_compar valeur ]	
<b>Les fonctions statistiques sur les DataFrame</b>	
mean()	valeur moyenne
min()	valeur minimale
count()	nombre d'éléments d'un DataFrame
sum()	la somme des valeurs
max()	valeur maximale

Manipulation des données d'un DataFrame	
<b>Modification du contenu d'une cellule :</b>	df.loc[N° ligne,"Nom_colonne"]=Valeur
<b>Modification des données d'une ligne :</b>	df.loc[N° ligne]=[Liste_Valeur]
<b>Ajout d'une colonne :</b>	df["Id_Colonne"]=valeur

Tri des éléments d'un DataFrame	
df_Trié = df.sort_index()	Tri par index
df_Trié = df.sort_index(axis =1)	Tri par colonne
<b>Tri des données</b>	
df_Trié = df.sort_values(by = [liste_colonnes], ascending = [True/False ] )	

Génération des graphes	
<b>Création d'un graphique à barres</b>	
df.plot.bar(x='Nom_Colonne_Abscisse', y='Nom_Colonne_Ordonnée', title='Titre_Graphique', color='couleur_Graphique')	
<b>Création d'un graphique en courbe</b>	
df.plot.line(x='Nom_Colonne_Abscisse', y='Nom_Colonne_Ordonnée', title='Titre_Graphique', color='couleur_Graphique')	
<b>Affichage d'un graphique</b>	
alias_matplotlib.show()	

Manipulation d'un DataFrame	
<b>Renommage des libellés d'une ou de plusieurs colonnes :</b>	
df.rename( columns = {"col1": "nouveau_nom_col1", ..., "colN": "nouveau_nom_colN"}, inplace=True False)	
<b>Suppression des lignes :</b>	
df=df.drop([Indices_Lignes])	
<b>Suppression des colonnes :</b>	
df=df.drop(columns=[Nom_Colonne1,...,Nom_ColonneN])	