

Aide-mémoire

Analyse de données avec Pandas

Importation de la bibliothèque Pandas

```
import pandas as pd
```

Importation de la bibliothèque matplotlib

```
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.Nom_colonne	Le contenu d'une colonne
df["Nom_colonne"]	Le contenu d'une colonne
df[["col1", "col2", ...]]	Le contenu de plusieurs colonnes
df.head(n)	Les n premières lignes
df.tail(n)	Les n dernières lignes

Les données des lignes :

```
df[ligne_initiale : ligne_finale]
```

Les données en utilisant les indices :

```
df.iloc[lg_i : lg_f , col_i : col_f]
```

Les données des cellules en utilisant des étiquettes :

```
df.loc[lg_i : lg_f , col_i : col_f]
```

Les données selon une condition :

```
df[ df ["nom_colonne"] op_compar valeur]
```

Les fonctions statistiques sur les DataFrame

mean() : valeur moyenne	sum() : la somme des valeurs
min() : valeur minimale	max() : valeur maximale
count() : nombre d'éléments d'un DataFrame	

Manipulation des données d'un DataFrame

Modification du contenu d'une cellule :

```
df.loc[N° ligne,"Nom_colonne"]=Valeur
```

Modification des données d'une ligne :

```
df.loc[N° ligne]=[Liste_Valeur]
```

Ajout d'une colonne :

```
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

Tri des données

```
df_Trié = df.sort_values(by = [liste_colonnes], ascending = [True/False ] )
```

Génération des graphes

Création d'un graphique à barres

```
df.plot.bar(x='Nom_Colonne_Abscisse', y='Nom_Colonne_Ordonnée', title='Titre_Graphique', color='couleur_Graphique')
```

Création d'un graphique en courbe

```
df.plot.line(x='Nom_Colonne_Abscisse', y='Nom_Colonne_Ordonnée', title='Titre_Graphique', color='couleur_Graphique')
```

Affichage d'un graphique

```
alias_matplotlib.show()
```

Manipulation d'un DataFrame

Renommage des libellés d'une ou de plusieurs colonnes :

```
df.rename( columns = {"col1":"nouveau_nom_col1",..., "colN":"nouveau_nom_colN"}, inplace=True | False)
```

Suppression des lignes :

```
df=df.drop([Indices_Lignes])
```

Suppression des colonnes :

```
df=df.drop (columns=["Nom_Colonne1",..., "Nom_ColonneN"])
```