Mémo Pandas

ACTION	COMMANDE
Lire un fichier CSV en créant le Dataframe T	T = pd.read_csv('nom.csv')
Importer le module pandas	import pandas as pd
Donner le nombre de lignes de T	len(T)
Donner la taille de T	T.shape
Donner le nom et type de chaque colonne	T.dtypes
Donner les 5 premières lignes de T	T.head()
Donner les 10 premières lignes de T	T.head(10)
Donner les 5 dernières lignes de T	T.tail()
Calculer les indicateurs statistiques de T	T.describe()
Calculer les indicateurs statistiques de la colonne C	T['C'].describe()
Calculer la moyenne de la colonne C	T['C'].mean()
Calculer l'écart-type de la colonne C	T['C'].std()
Calculer la médiane de la colonne C	T['C'].median()
Calculer la somme de la colonne C	T['C'].sum()
Calculer l'effectif de la colonne C	T['C'].count()
Donner les effectifs de chaque valeur de la colonne C	T['C'].value_counts()
Trier T par ordre croissant selon la colonne C	T.sort_values(by='C')
Trier T par ordre décroissant selon la colonne C	T.sort_values(by='C', ascending=False)
Extraire la colonne C	T['C']
Extraire les colonnes C1 et C2	T[['C1', 'C2']]
Filtrer les lignes avec les valeurs de C non nulle	T.query('C != 0')
Filtrer les lignes avec A > 2 et B < 6	T.query('A > 2 & B < 6')
Filtrer les lignes avec A < 2 ou B > 6	T.query('A < 2 | B > 6')
Filtrer les lignes de la colonne C correspondant à 'foo'	T.query(' C == "foo"')
Filtrer les lignes de la colonne C contenant 'foo'	T.query('C.str.contains("pu")')
Filtrer les lignes de la colonne C avec les éléments d'une liste L	T.query('C in @L')
Échantillon aléatoire de 100 lignes	T.sample(n=100)
Filtrer les lignes d'index 7 et 8	T.query('index in [7,8]')
Supprimer les lignes en double selon une liste de noms de colonne	T.drop_duplicates(subset=[c1, c2, …])

Grouper et compter selon la colonne C	T.groupby('C').count()
Grouper et calculer la moyenne selon la colonne C	T.groupby('C').mean()
Grouper et calculer la somme selon la colonne C	T.groupby('C').sum()
Grouper et trouver le max selon la colonne C	T.groupby('C').max()
Créer la colonne C comme somme terme de colonne A et de la colonne B	T['C'] = T['A'] + T['B']
Créer la colonne C comme produit de chaque terme de la colonne A par 3	T['C'] = 3 * T['A']
Créer la colonne C comme la somme de chaque time de la colonne A par 4	T['C'] = T['A'] + 4
Appliquer une fonction f à la colonne A	T['C'] = T['A'].apply(f)