

cheatsheetPandas

November 16, 2024

Importer Pandas

```
import pandas as pd
```

Créer un dataframe manuellement

```
df = pd.DataFrame({  
    'Colonne1': [1, 2, 3],  
    'Colonne2': ['A', 'B', 'C']  
})
```

Créer un dataframe via un fichier CSV

```
# Le fichier doit se trouver dans le même dossier que le fichier de code .ipynb  
df = pd.read_csv('fichier.csv')
```

Aperçu des données

- Afficher les premières lignes :

```
df.head() # Par défaut, affiche les 5 premières lignes  
df.head(10) # Affichera les 10 premières lignes
```

- Afficher les dernières lignes :

```
df.tail() # Par défaut, affiche les 5 dernières lignes  
df.tail(10) # Affichera les 10 dernières lignes
```

- Informations sur les données :

```
df.info() # Infos générales sur le DataFrame
```

- Statistiques descriptives :

```
df.describe()
```

Sélection de données

- Sélectionner une colonne :

```
df_selection = df['Colonne1']
```

- Sélectionner plusieurs colonnes :

```
df_selection = df[['Colonne1', 'Colonne2']]
```

Filtrage de données

```
df_filtre = df[df['Colonne1'] > valeur1]
df_filtre = df[(df['Colonne1'] > 1) & (df['Colonne2'] < valeur2)]
```

Manipulation des données

- Ajouter une colonne :

```
df['NouvelleColonne'] = df['Colonne1'] * 2
```

- Renommer une colonne :

```
df.rename(columns={'Colonne1': 'NouveauNom'}, inplace=True) # inplace permet de modifier i
```

- Renommer plus d'une colonne :

```
liste1 = ['nom1', 'nom2', ...,] # Liste des anciens noms
liste2 = ['nom1', 'nom2', ...,] # Liste des anc nouveaux noms
```

```
for i in range(len(liste1)):
    df = df.rename(columns={liste1[i]: liste2[i]})
```

- Supprimer une colonne :

```
df = df.drop(columns=['Colonne2'], inplace=True)
```

- Supprimer plus d'une colonne :

```
df = df.drop(['Colonne1', 'Colonne2'], axis=1)
```

Gestion des valeurs manquantes

- Détecter les valeurs manquantes :

```
df.isna() # Indique True si la valeur est manquante, False sinon
df.isna().sum() # Affiche le nombre de valeurs manquantes par colonne
```

Opérations sur les données

- Tri des données :

```
df.sort_values(by='Colonne1', ascending=False)
```

- Compter les données :

```
nombre = df['Colonne'].value_counts()
```

- Convertir un DataFrame en liste:

```
liste = df.values.tolist()
```

- Convertir une liste en DataFrame:

```
df = pd.DataFrame(liste, columns=df.columns)
```