

Python : Pandas

Achref El Mouelhi

Docteur de l'université d'Aix-Marseille
Chercheur en programmation par contrainte (IA)
Ingénieur en génie logiciel

elmouelhi.achref@gmail.com



Plan

1 Introduction

2 Installations

3 Lecture

- Lecture intégrale
- Lecture ligne par ligne
- Lecture de colonnes par nom
- Lecture de colonnes par index
- Lecture de toutes les feuilles

4 Écriture

- Nouveau classeur et nouvelle feuille
- Classeur existant et nouvelle feuille
- Classeur et feuille existants
- Formule

Python

Pandas

- bibliothèque open-source pour le langage de programmation **Python**
- offrant des structures de données (unidimensionnelle comme `Series` et bidimensionnelles comme `DataFrame`) et des outils de manipulation de données
- permettant d'importer/exporter des données à partir/vers de diverses sources, telles que **CSV, Excel, JSON...**
- largement utilisé dans le domaine de la science des données et de l'analyse de données pour manipuler, nettoyer, analyser et visualiser des ensembles de données de manière efficace

Python

Documentation officielle

<https://pandas.pydata.org/>

Python

Démarche

- Créez un répertoire cours-pandas dans votre espace de travail
- Lancez **VSC** et allez dans File > Open Folder... et choisissez cours-pandas
- Dans cours-pandas, créez un fichier main.py

Python

Pour installer Pandas, lancez la commande

```
pip install pandas openpyxl
```

Python

Pour installer Pandas, lancez la commande

```
pip install pandas openpyxl
```

Démarche

- Openpyxl : bibliothèque **Python** open-source qui permet de travailler avec des fichiers **Excel** au format `.xlsx`
Documentation officielle : <https://openpyxl.readthedocs.io/>
- ezodf : bibliothèque **Python** qui permet de lire et de modifier des fichiers **Excel** ayant l'extension `.ods`

Python

Pour utiliser Pandas, il faut l'importer (ici sous l'alias pd)

```
import pandas as pd
```

Python

Pour utiliser Pandas, il faut l'importer (ici sous l'alias pd)

```
import pandas as pd
```

Télécharger et placer le fichier Excel à la racine du projet

<http://elmouelhia.free.fr/fichier.xlsx>

Python

Pour utiliser Pandas, il faut l'importer (ici sous l'alias pd)

```
import pandas as pd
```

Télécharger et placer le fichier Excel à la racine du projet

<http://elmouelhia.free.fr/fichier.xlsx>

Créons une variable pour le fichier Excel téléchargé

```
filename = 'fichier.xlsx'
```

Python

Pour lire la feuille enseignants du fichier Excel (df : DataFrame)

```
df = pd.read_excel(filename, 'enseignants')
```

Python

Pour lire la feuille enseignants du fichier Excel (df : DataFrame)

```
df = pd.read_excel(filename, 'enseignants')
```

Pour afficher le contenu de la feuille enseignants

```
print(df.head())
```

Python

Pour lire la feuille enseignants du fichier Excel ligne par ligne

```
for ligne in df.iterrows():
    print(ligne)
```

Python

Pour récupérer l'ensemble de colonnes (les noms)

```
for nom_colonne in df.columns:  
    print(nom_colonne)
```

© Achref EL MOUELHIFI

Python

Pour récupérer l'ensemble de colonnes (les noms)

```
for nom_colonne in df.columns:  
    print(nom_colonne)
```

Pour itérer sur chaque élément de la colonne

```
for nom_colonne in df.columns:  
    colonne = df[nom_colonne]  
  
    for valeur in colonne:  
        print(valeur)
```

Python

Pour récupérer l'ensemble de colonnes (les indices)

```
for index_colonne in range(len(df.columns)):  
    colonne = df.iloc[:, index_colonne]
```

© Achref EL MOUELHID

Python

Pour récupérer l'ensemble de colonnes (les indices)

```
for index_colonne in range(len(df.columns)):  
    colonne = df.iloc[:, index_colonne]
```

Pour itérer sur chaque élément de la colonne selon l'indice

```
for index_colonne in range(len(df.columns)):  
    colonne = df.iloc[:, index_colonne]  
  
    for valeur in colonne:  
        print(valeur)
```

Python

Pour lire toutes les feuilles d'un document Excel, on affecte la valeur `None` à `sheet_name`, la valeur de retour de `read_excel` sera un dictionnaire

```
all_sheets = pd.read_excel(filename, sheet_name=None)
```

Python

Pour lire toutes les feuilles d'un document Excel, on affecte la valeur `None` à `sheet_name`, la valeur de retour de `read_excel` sera un dictionnaire

```
all_sheets = pd.read_excel(filename, sheet_name=None)
```

Pour afficher le contenu de toutes les feuilles

```
for nom_feuille, df in all_sheets.items():
    print(f"Feuille : {nom_feuille}")
    print(df)  # affiche les données de la feuille
    print("\n")  # ajoute une ligne vide pour la lisibilité
```

Python

Commençons par importer Workbook **de** openpyxl

```
from openpyxl import Workbook
```

© Achref EL MOUELHI ©

Python

Commençons par importer Workbook **de** openpyxl

```
from openpyxl import Workbook
```

Pour créer un nouveau classeur Excel

```
classeur = Workbook()
```

Python

Commençons par importer Workbook **de openpyxl**

```
from openpyxl import Workbook
```

Pour créer un nouveau classeur Excel

```
classeur = Workbook()
```

Pour créer une nouvelle feuille de calcul et la nommer MaFeuille

```
feuille = classeur.active  
feuille.title = "MaFeuille"
```

Python

Pour ajouter des données dans la feuille de calcul

```
feuille['A1'] = 'Nom'  
feuille['B1'] = 'Age'  
  
feuille['A2'] = 'Wick'  
feuille['B2'] = 45  
  
feuille['A3'] = 'Dalton'  
feuille['B3'] = 55
```

Python

Pour ajouter des données dans la feuille de calcul

```
feuille['A1'] = 'Nom'  
feuille['B1'] = 'Age'  
  
feuille['A2'] = 'Wick'  
feuille['B2'] = 45  
  
feuille['A3'] = 'Dalton'  
feuille['B3'] = 55
```

Pour enregistrer le classeur Excel dans un fichier

```
filename = 'file.xlsx'  
classeur.save(filename)  
  
print(f'Le fichier "{filename}" a été créé avec succès.')
```

Python

Commençons par importer `load_workbook` **de** `openpyxl`

```
from openpyxl import load_workbook
```

Python

Commençons par importer `load_workbook` **de** `openpyxl`

```
from openpyxl import load_workbook
```

Pour charger le classeur existant

```
filename = 'file.xlsx'  
classeur = load_workbook(filename)
```

Python

Commençons par importer `load_workbook` **de** `openpyxl`

```
from openpyxl import load_workbook
```

Pour charger le classeur existant

```
filename = 'file.xlsx'  
classeur = load_workbook(filename)
```

Pour ajouter une nouvelle feuille au classeur et la nommer `EncoreUne`

```
nouvelle_feuille = classeur.create_sheet("EncoreUne")
```

Python

Pour ajouter des données dans la feuille de calcul

```
nouvelle_feuille['A1'] = 'Prenom'  
nouvelle_feuille['B1'] = 'Salaire'  
  
nouvelle_feuille['A2'] = 'John'  
nouvelle_feuille['B2'] = 3000  
  
nouvelle_feuille['A3'] = 'Jack'  
nouvelle_feuille['B3'] = 2500
```



Python

Pour ajouter des données dans la feuille de calcul

```
nouvelle_feuille['A1'] = 'Prenom'  
nouvelle_feuille['B1'] = 'Salaire'  
  
nouvelle_feuille['A2'] = 'John'  
nouvelle_feuille['B2'] = 3000  
  
nouvelle_feuille['A3'] = 'Jack'  
nouvelle_feuille['B3'] = 2500
```



Pour sauvegarder les changements dans le fichier existant

```
classeur.save(filename)
```

```
print(f'Une nouvelle feuille a été ajoutée dans "{filename}" .')
```

Python

Commençons par importer `load_workbook` **de** `openpyxl`

```
from openpyxl import load_workbook
```

Python

Commençons par importer `load_workbook` **de** `openpyxl`

```
from openpyxl import load_workbook
```

Pour charger le classeur existant

```
filename = 'file.xlsx'  
classeur = load_workbook(filename)
```

Python

Commençons par importer `load_workbook` **de** `openpyxl`

```
from openpyxl import load_workbook
```

Pour charger le classeur existant

```
filename = 'file.xlsx'  
classeur = load_workbook(filename)
```

Pour ouvrir une feuille existante

```
feuille = classeur["EncoreUne"]
```

Python

Pour modifier ou ajouter des données dans la feuille

```
feuille['A1'] = 'Firstname'  
feuille['B1'] = 'Salary'  
feuille['A4'] = 'Bill'  
feuille['B4'] = 10000
```

Python

Pour modifier ou ajouter des données dans la feuille

```
feuille['A1'] = 'Firstname'  
feuille['B1'] = 'Salary'  
feuille['A4'] = 'Bill'  
feuille['B4'] = 10000
```

Pour sauvegarder les changements dans le fichier

```
classeur.save(filename)
```

```
print(f'La feuille "{feuille.title}" a été modifiée avec succès.')
```

Python

Pour associer une formule à une cellule

```
from openpyxl import load_workbook

filename = 'file.xlsx'
classeur = load_workbook(filename)
feuille = classeur["EncoreUne"]

feuille['A6'] = 'Total'
feuille['B6'] = '=SUM(B2:B4)'

classeur.save(filename)

print('La formule a été ajoutée à la cellule B6.')
```