

Dev-Data Pratique : Essentiel Pandas

À rendre avant le Lundi 02-06-2025 à 7H59min

Objectifs pédagogiques

- Comprendre les structures de données Series et DataFrame
- Savoir importer, explorer, filtrer et transformer des données tabulaires
- Appliquer les fonctions d'agrégation, de nettoyage et de jointure

Partie 1 – Création et lecture de DataFrames

1. Créer une Series manuellement et l'afficher
2. Créer un DataFrame à partir d'un dictionnaire Python
3. Lire un fichier CSV avec `pd.read_csv()`
4. Afficher les 5 premières et dernières lignes (`head()`, `tail()`)
5. Afficher les informations générales (`info()`, `shape`, `dtypes`)

Partie 2 – Accès aux données et filtrage

1. Accéder à une colonne (`df["col"]`) ou plusieurs (`df[["col1", "col2"]]`)
2. Accéder à une ligne par index (`loc`, `iloc`)
3. Filtrer les lignes avec une condition logique (`df[df["col"] > valeur]`)
4. Appliquer plusieurs conditions combinées (`&`, `|`)

Partie 3 – Nettoyage des données

1. Détecter et supprimer les doublons (`drop_duplicates()`)
2. Gérer les valeurs manquantes : `isna()`, `fillna()`, `dropna()`
3. Modifier les types de colonnes (`astype()`)
4. Renommer les colonnes (`rename()`)

Partie 4 – Transformation et calculs

1. Créer une nouvelle colonne à partir d'une opération entre colonnes
2. Appliquer une fonction à une colonne (`apply()`, `lambda`)
3. Grouper les données par catégorie (`groupby()` + `sum()`, `mean()`...)
4. Trier les données par valeurs (`sort_values()`)

Partie 5 – Fusion, jointure et reshape

1. Fusionner deux DataFrames avec `merge()`
2. Concaténer avec `concat()` (horizontale et verticale)

3. Transformer les données avec `pivot()`, `melt()`

Partie 6 – Mini-challenge Pandas

Contexte : Analyse d'un jeu de données de ventes (clients, produits, commandes).

- Calculer le chiffre d'affaires total et par client
- Afficher les 5 meilleurs clients
- Calculer le panier moyen par commande
- Détecter les produits les moins vendus
- Produire un DataFrame final prêt à être visualisé

Livrable attendu

- Un **Notebook Jupyter** propre avec :
 - Les codes bien organisés
 - Des explications Markdown
 - Un récapitulatif des difficultés rencontrées et des apprentissages

Bibliothèque obligatoire :

```
import pandas as pd
```

NB : Utilisez les fichiers et/ou données produit lors des modules précédant