Data analyse avec Pandas et Numpy

Sommaire

Sujet : Pandas
Catégorie : frameworks
Auteur(s) : BELHADJ Karim
Version : 1.0

Table des matières

I. Chapitre 2 Présentation des Series 1

Comparaison entre structures ndarray et Series 1
Méthodes de création d'une `Series` 1
Attributs et méthodes d'inspection d'une `Series` 1
Les `Series` temporelles : gestion des dates et du temps 1
La conversion de type dans une `Series` (`.astype()`) 1
Exemple de synthèse 1
Exercices de fin de chapitre 1

II. Chapitre 3 Manipuler les Series 1

Opérations arithmétiques et alignement automatique 1
Extraction et recherche de données 1
Travailler avec les index d'une `Series` 1
Transformation des valeurs d'une `Series` 1
Exemple de synthèse 1
Exercices de fin de chapitre 1

III. Chapitre 4 Présentation un DataFrame 1

Series vs DataFrame (et comparaison avec NumPy) 1
Organisation interne et propriétés d'un DataFrame 1
Gestion des données invalides (`NaN`) 1
Créations multiples : listes, tuples et dictionnaires 1
Chargement de fichiers externes : CSV et Excel 1
Exemple de synthèse 1
Exercices de fin de chapitre 1

IV. Chapitre 5 Manipuler un DataFrame 1

Travailler sur les colonnes : sélection, calcul et nettoyage 1
Travailler sur les lignes : filtrage, ajout et suppression 1
Travailler sur les labels et index 1
Extraction de données : `.loc` vs `.iloc` 1
Exemple de synthèse 1
Exercices de fin de chapitre 1

V. Chapitre 6 Opérations sur les données 1

Opération de Tri 1
Itération sur les lignes et modification 1

Data analyse avec Pandas et Numpy

Sommaire

Utilisation des fonctions `apply` 2
Utilisation des `pipe` 2
Recherches et critères de filtrage 2
Exemple de synthèse 2
Exercices de fin de chapitre 2

VI. Chapitre 7 Filtrage et Mappage dans les DataFrames Pandas 2

Le filtrage conditionnel de lignes (Indexation booléenne) 2
La méthode `.filter()` : Sélectionner des colonnes ou des index par nom 2
Le Mappage (`.map()`) : Transformer et traduire des valeurs 2
Traduire des codes catégories 2
Aller plus loin : `.apply()` et `.applymap()` (ou `.map()` global) 2
Exemple de synthèse 2
Exercices de fin de chapitre 2

VII. Chapitre 8 Nettoyage et Préparation des Données 2

Détecter les données invalides (`NaN`, `isna()`, `notna()`) 2
Traiter les valeurs manquantes : `fillna()`, `dropna()`, `interpolate()` 2
Détecter et supprimer les doublons : `duplicated()`, `drop_duplicates()` 2
Repérer les valeurs aberrantes (Outliers) 2
Nettoyer les chaînes de caractères avec l'accesseur `.str` 2
Le type `category` : gain de mémoire et de performance 2
Exemple de synthèse : 2
Exercices de fin de chapitre 2

VIII. Chapitre 9 Gestion des dates et séries temporelles 2

Les 4 modèles fondamentaux du temps dans Pandas 2
Manipulation de plages temporelles avec `period_range` 2
Combiner `pd.date_range()` avec un dictionnaire de données 2
Modèles de temps 2
Manipulation de plages temporelles 2
Extraire le mois, le jour et l'heure 2
Exemple de synthèse 2
Exercices de fin de chapitre 2

IX. Chapitre 10 Statistiques et agrégations 2

Statistiques descriptives 2
Agrégations par catégories (`groupby`) 2
Agrégation avec plusieurs fonctions de réduction (`.agg`) 2
Comparer `.agg` vs `.groupby` 2
Exemple de synthèse 2
Exercices de fin de chapitre 2

Data analyse avec Pandas et Numpy

Sommaire

X. Chapitre 11 Jointures et fusion de données 3

- Concaténation de structures 3
- pd.merge() : Le couteau suisse des jointures (Basé sur des colonnes ou des clés) 3
- La méthode `.join()` : Le spécialiste des index 3
- Exemple de synthèse 3
- Exercices de fin de chapitre 3

XI. Chapitre 12 Persistance et formats 3

- Le format CSV (*Comma-Separated Values*) : Le standard tabulaire 3
- Le format JSON (*JavaScript Object Notation*) : Le roi du Web 3
- Le format XML (*Extensible Markup Language*) : Les données semi-structurées 3
- Accès aux données stockées en base relationnelle (`sql`) 3
- Persistance des données via le format Pickle 3
- Synthèse des méthodes de persistance 3
- Exemple de synthèse 3
- Exercices de fin de chapitre 3

XII. Chapitre 13 Gestion des structures multidimensionnelles 3

- Introduction aux Panel Data 3
- Utilisation du `MultiIndex` 3
- Sélection et tri sur un `MultiIndex` 3
- Tableaux croisés : `pivot()`, `pivot_table()` et `crosstab()` 3
- Remodeler les données : `stack()`, `unstack()` et `melt()` 3
- Exemple de synthèse : Analyse multidimensionnelle e-commerce 3
- Exercices de fin de chapitre 3

XIII. Chapitre 14 Rappel Pyplot - Les fondations graphiques de Python 3

- Architecture de Matplotlib : `Figure` et `Axes` 3
- Tracer un graphique simple : `plt.plot()` 3
- Titres, étiquettes, légendes et styles 3
- Plusieurs graphiques dans une figure : `plt.subplots()` 3
- Enregistrer une figure 3
- Exemple de synthèse 3
- Exercices de fin de chapitre 3

XIV. Chapitre 15 Visualisation de données avec Pandas (`df.plot()`) 3

- Visualisation avec Matplotlib, Pyplot et Seaborn 3
- Le catalogue des représentations graphiques 3
- Habillage, personnalisation et multi-graphiques 3
- Exemple de synthèse 3
- Exercices de fin de chapitre 3

Data analyse avec Pandas et Numpy

Sommaire

XV. Chapitre 16 Étude de cas 4

Présentation du jeu de données et des questions métier 4
Chargement et inspection 4
Nettoyage et préparation 4
Agrégations et croisements 4
Visualisation et restitution des résultats 4
Exemple de synthèse 4
Exercices de fin de chapitre 4

XVI. Chapitre 17 Cheatsheet NumPy / Pandas / Matplotlib 4

1. NumPy : le socle 4
2. Series 4
3. DataFrame : création et structure 4
4. Colonnes 4
5. Lignes et filtrage 4
6. Index, `.loc` et `.iloc` 4
7. Tri, itération, apply, pipe 4
8. Nettoyage 4
9. Dates et temps 4
10. Statistiques et `groupby` 4
11. Fusion de données 4
12. Structures multidimensionnelles 4
13. Export / persistance 4
14. Visualisation 4
15. Pipeline type d'analyse 4

Pièges à retenir 4