Pandas Avancé

Lien : https://innov-systems.com/formation/pandas-avance

DURÉE
4 jours (28h)

RÉFÉRENCE
BSI423

CATÉGORIE
Big Data : développement, langages et NOSQL

OBJECTIFS DE LA FORMATION

À l'issue de cette formation, vous serez capable de :

- Approfondir la maîtrise de la librairie Pandas pour réaliser des analyses de données complexes et optimisées
- Mettre en œuvre des techniques avancées de manipulation, d’agrégation et de transformation des données
- Optimiser les performances des traitements avec Pandas et ses librairies associées
- Appliquer les bonnes pratiques de structuration et de nettoyage des données pour la Data Science

POUR QUI ?

- Data scientists, analystes de données, statisticiens
- Développeurs Python amenés à manipuler de grandes volumétries de données
- Responsables d’études et décideurs techniques impliqués dans l’exploitation des données

Innov Systems

Programme détaillé

1 / Rappels fondamentaux et bonnes pratiques avec Pandas

- Manipulation avancée des DataFrame et Series (sélection, indexation, slicing)
- Import/export optimisé (CSV, Excel, SQL, Parquet, JSON) et gestion des encodages
- Nettoyage, formatage et typage des données (gestion des NaN, conversions, optimisations mémoire)

2 / Gestion avancée des dates, chaînes et types catégoriels

- Manipulation de dates et heures (datetime, resampling, time zones)
- Traitement avancé des chaînes de caractères avec `.str` et regex
- Utilisation des types catégoriels pour améliorer la performance et réduire la mémoire

3 / Agrégations et GroupBy avancés

- GroupBy multi-index et multi-clés avec fonctions d'agrégation personnalisées
- Utilisation combinée de `apply`, `agg` et `transform` pour des traitements complexes
- Création de métriques sur mesure pour des analyses ciblées

4 / Tableaux croisés et pivots dynamiques

- Création et personnalisation de tables pivots pour la synthèse des données
- Tableaux croisés et matrices de contingence pour l'analyse relationnelle
- Pivot et unpivot (melt) pour remodeler les données

5 / Jointures, fusions et concaténations complexes

- Fusion de plusieurs DataFrames avec différentes stratégies (`merge`, `join`, `concat`)
- Gestion des clés multiples et jointures hiérarchiques
- Détection et résolution des doublons après fusion

Innov Systems

6 / Optimisation des performances et parallélisation

- Vectorisation et élimination des boucles pour accélérer les traitements
- Intégration de NumPy pour les opérations bas niveau
- Utilisation de Modin et Dask pour le traitement parallèle sur gros volumes
- Accélération avec Numba et Cython pour les calculs intensifs

7 / Visualisation avancée avec Pandas et intégration Seaborn/Matplotlib

- Création de graphiques directement depuis Pandas (plots, hist, boxplots)
- Combinaison avec Seaborn pour des visualisations statistiques avancées
- Optimisation des graphiques pour un rendu professionnel

8 / Atelier de mise en pratique intégral

- Mise en place d’un workflow complet d’analyse de données réelles
- Nettoyage, transformation, agrégation et visualisation sur un projet complet
- Application des bonnes pratiques de performance et de structuration des données

Approche pédagogique

- Support Ecrit et Projection
- Exposés Interactifs, Podcasts et Vidéos
- Brainstorming et Jeux de Rôle
- Cas Pratiques et Labs inclus pour leur impact opérationnel
- Test de Validation des Acquis des Connaissances

Innov Systems

Prochaines dates programmées

17 au 20 Nov. 2026	Distanciel
09 au 12 Fév. 2027	Présentiel - Casablanca
06 au 09 Avr. 2027	Présentiel - Casablanca
01 au 04 Juin 2027	Présentiel - Casablanca
27 au 30 Juil. 2027	Présentiel - Casablanca

Autres dates possibles sur demande. Contactez-nous pour organiser une session intra-entreprise.

Réservation & Renseignements

Téléphone : +212 522 247 210
Email : contact@innov-systems.com
Web : https//www.innov-systems.com

Innov Systems — Tous droits réservés

Innov Systems