

Pandas avancé

La bibliothèque Python pour l'analyse de données

Cours Pratique de 2 jours - 14h
Réf : PND - Prix 2024 : 1 450€ HT

Vous avez découvert lors de vos apprentissages Python, Pandas : la bibliothèque pour l'analyse de données. Vous souhaitez être guidé, par un expert du domaine, pour approfondir l'exploration de cette librairie nécessaire à tous projets de data science, Cette formation est faite pour vous.

OBJECTIFS PÉDAGOGIQUES

À l'issue de la formation l'apprenant sera en mesure de :

- Maîtriser la librairie Pandas pour l'analyse de données
- Connaître les subtilités des groupbys
- Savoir manipuler les tables pivots et les tableaux croisés
- Acquérir des notions sur l'accélération des calculs avec Pandas
- Connaître les bonnes pratiques en Data Science

LE PROGRAMME

dernière mise à jour : 05/2022

1) Tour d'horizon de la librairie Pandas

- Rappels sur les notions de bases de Pandas.
- Lecture de fichiers de données (csv, excel, SQL, parquet).
- Description du jeu de données, et analyse statistique simple.
- Implémenter des analyses et des visualisations différentes en fonction du type de données.
- Gestion des données manquantes.
- Manipulation de dates pour les Time Series.
- Gestion des chaînes de caractères.
- Mise en place des bonnes pratiques en Data Science.

Travaux pratiques : Mise en place d'un environnement virtuel pour la Data Science, lecture d'un fichier csv et xls, brève analyse statistique et description des jeux de données.

2) Maîtriser les subtilités des groupbys

- Les groupbys pour appréhender les modalités dans les jeux de données.
- Groupby à simple indice avec les fonctions d'agrégations classiques.
- Personnalisation des fonctions d'agrégations.
- Groupby à multiples indices.
- Différence entre les fonctions apply et transform.
- Rappels sur les fonctions anonymes.

Travaux pratiques : Sur 2 jeux de données économiques, mise en pratique du groupby et visualisation des données. Création d'un toy dataset et utilisation du groupby.

3) Tables pivots et tableaux croisés

- Fonctions d'agrégation et tables pivots.
- Matrice de contingence.

PARTICIPANTS

Développeurs, ingénieurs et toute personne analysant les données avec des compétences de développement.

PRÉREQUIS

Maîtrise de Python

COMPÉTENCES DU FORMATEUR

Les experts qui animent la formation sont des spécialistes des matières abordées. Ils ont été validés par nos équipes pédagogiques tant sur le plan des connaissances métiers que sur celui de la pédagogie, et ce pour chaque cours qu'ils enseignent. Ils ont au minimum cinq à dix années d'expérience dans leur domaine et occupent ou ont occupé des postes à responsabilité en entreprise.

MODALITÉS D'ÉVALUATION

Le formateur évalue la progression pédagogique du participant tout au long de la formation au moyen de QCM, mises en situation, travaux pratiques...
Le participant complète également un test de positionnement en amont et en aval pour valider les compétences acquises.

MOYENS PÉDAGOGIQUES ET TECHNIQUES

- Les moyens pédagogiques et les méthodes d'enseignement utilisés sont principalement : aides audiovisuelles, documentation et support de cours, exercices pratiques d'application et corrigés des exercices pour les stages pratiques, études de cas ou présentation de cas réels pour les séminaires de formation.
- À l'issue de chaque stage ou séminaire, ORSYS fournit aux participants un questionnaire d'évaluation du cours qui est ensuite analysé par nos équipes pédagogiques.
- Une feuille d'émargement par demi-journée de présence est fournie en fin de formation ainsi qu'une attestation de fin de formation si le stagiaire a bien assisté à la totalité de la session.

MODALITÉS ET DÉLAIS D'ACCÈS

L'inscription doit être finalisée 24 heures avant le début de la formation.

ACCESSIBILITÉ AUX PERSONNES HANDICAPÉES

Vous avez un besoin spécifique d'accessibilité ? Contactez Mme FOSSE, référente handicap, à l'adresse suivante psh-accueil@orsys.fr pour étudier au mieux votre demande et sa faisabilité.

- Tableaux croisés.

Travaux pratiques : Sur 2 jeux de données économiques, mise en pratique des tables pivots et tableaux croisés.

4) Jointure de tables

- Notions d'axes.

- Concaténation.

- Merge selon une ou plusieurs clés.

- Jointure par rapport aux indices.

Travaux pratiques : Sur 2 jeux de données économiques, mise en pratique des différents types de jointures.

5) Accélération du calcul avec Pandas

- Boucler sur les lignes et les colonnes.

- Revenir aux basiques avec NumPy.

- Exemples de multiprocessing avec la librairie Modin.

- Exemples de multiprocessing avec la librairie Numba.

Travaux pratiques : Sur un jeu de données volumineux, mise en pratique des différentes notions abordées lors du cours.

LES DATES

CLASSE À DISTANCE

2024 : 18 juil., 07 nov.

PARIS

2024 : 11 juil., 24 oct.