

FORMATION & CERTIFICATION • Analyste de données certifié Python

# Analyste de données certifié Python – Niveau Associé

Formation disponible en présentiel ou à distance à la demande. Maîtrisez Python et les principaux outils d'analyse de données (NumPy, pandas, visualisation, statistiques) pour exploiter et valoriser les données dans un contexte professionnel.

Formation présentiel & distanciel

Centre d'examen Pearson VUE

Formateurs certifiés Microsoft

**5 jours**

30–35 h

**Examen**

Analyste de données certifié Python

**Modalité**

Présentiel  
—  
Distanciel

**Niveau**

Intermédiaire

**INSCRIPTION / RÉSERVATION**



Je m'inscris maintenant



## OBJECTIFS PÉDAGOGIQUES

- Acquérir une maîtrise opérationnelle de Python appliqué à l'analyse de données.
- Manipuler des données tabulaires avec pandas et NumPy (chargement, filtrage, jointures, agrégations).
- Réaliser une analyse exploratoire complète (statistiques descriptives, visualisations, corrélations).
- Mettre en œuvre des modèles de base d'apprentissage automatique pour la prédiction et la classification.
- Valider les compétences à travers un projet fil rouge et un examen de certification.

## PUBLIC CIBLE

- Professionnels souhaitant se spécialiser dans l'analyse de données avec Python.
- Ingénieurs, statisticiens, économistes, responsables métiers amenés à exploiter des données.
- Étudiants et personnes en reconversion vers les métiers de la Data (Data Analyst, Business Analyst...).
- Toute personne ayant déjà des bases en Python et souhaitant aller plus loin dans l'analyse de données.

## PRÉREQUIS

- Connaissances de base en programmation (variables, boucles, conditions, fonctions).
- Notions de base en statistiques (moyenne, médiane, écart-type) souhaitables.
- Une première expérience avec Python ou avoir suivi une formation de niveau débutant (PCEP ou équivalent).

## PROGRAMME DE LA FORMATION – DÉTAILLÉ

### **Introduction à l'analyse de données avec Python**

#### **Contexte et concepts clés**

- Rôle de l'analyste de données et panorama des métiers de la Data.
- Cycle de vie d'un projet d'analyse de données : de la collecte à la mise en valeur.
- Présentation de l'écosystème Python pour la Data (NumPy, pandas, Matplotlib, Seaborn, scikit-learn...).

#### **Environnement de travail**

- Installation et prise en main de Python, Jupyter Notebook et/ou VS Code.
- Organisation d'un projet : structure de dossiers, gestion des dépendances (pip, virtualenv).
- Bonnes pratiques de base : commentaires, lisibilité, sauvegarde des notebooks.

### **Rappels et approfondissement Python pour la Data**

#### **Fondamentaux Python appliqués**

- Types de données (numériques, booléens, chaînes, listes, tuples, dictionnaires, ensembles).
- Contrôle de flux : conditions, boucles, compréhensions de listes.
- Fonctions, arguments, retour de valeurs et gestion des erreurs (exceptions).

#### **Manipulation de fichiers**

- Lecture et écriture de fichiers texte et CSV.
- Introduction aux formats JSON et aux API simples.
- Bonnes pratiques pour la manipulation de chemins et d'encodage.

### **NumPy et pandas : manipulation avancée de données**

#### **NumPy pour le calcul scientifique**

- Tableaux multidimensionnels (ndarray) et opérations vectorisées.
- Indexation, slicing, filtrage et masques booléens.
- Fonctions mathématiques et statistiques intégrées.

#### **pandas pour les données tabulaires**

- Objets Series et DataFrame : création, inspection et sélection.
- Chargement de données (CSV, Excel, JSON, bases de données).
- Sélection de colonnes/lignes, filtrage conditionnel, tri et réindexation.
- Ajout, suppression et renommage de colonnes.

## PROGRAMME DE LA FORMATION – DÉTAILLÉ

### **Nettoyage et préparation des données**

#### **Qualité des données**

- Identifier les valeurs manquantes, aberrantes et dupliquées.
- Traitement des valeurs manquantes : suppression, imputation simple, stratégies avancées.
- Gestion des doublons et cohérence des formats.

#### **Transformation des données**

- Opérations de mapping, regroupement de catégories, création de variables dérivées.
- Normalisation et standardisation des données numériques.
- Encodage des variables catégorielles (one-hot encoding, label encoding).
- Gestion des dates et temps : conversion, extraction de composantes (année, mois, jour...).

#### **Jointures et agrégations**

- Fusion de DataFrames (merge, join, concat).
- GroupBy, agrégations, tableaux croisés dynamiques.
- Cas pratiques : construction d'indicateurs métiers à partir de données brutes.

### **Analyse exploratoire et visualisation de données**

#### **Statistiques descriptives**

- Mesures de tendance centrale : moyenne, médiane, mode.
- Mesures de dispersion : variance, écart-type, quartiles.
- Distribution, asymétrie, kurtosis et détection des valeurs extrêmes.

#### **Visualisation avec Matplotlib et Seaborn**

- Graphiques de base : courbes, barres, secteurs, histogrammes.
- Visualisations avancées : boxplots, violin plots, heatmaps, pair plots.
- Personnalisation : titres, légendes, annotations, palettes de couleurs.
- Visualisation de corrélations et interprétation.

#### **Mise en forme des analyses**

- Structurer un notebook pour raconter une histoire avec les données (Data Storytelling).
- Export de graphiques et de rapports (PDF, images, Excel).

## PROGRAMME DE LA FORMATION – DÉTAILLÉ

### **Introduction au Machine Learning pour l'analyse de données**

#### **Concepts fondamentaux**

- Apprentissage supervisé vs non supervisé.
- Jeux d'entraînement, de validation et de test.
- Sur-apprentissage, sous-apprentissage, biais et variance.

#### **Prise en main de scikit-learn**

- Pipeline de traitement : préparation des données, entraînement, évaluation.
- Régression linéaire et régression logistique.
- Algorithmes de classification de base (k-NN, arbres de décision).
- Métriques d'évaluation : RMSE, précision, rappel, F1-score, matrice de confusion.

#### **Clustering et segmentation**

- Introduction au clustering (k-means) pour segmenter des données.
- Interprétation des clusters et illustrations sur des cas métiers simples.

### **Projet fil rouge et bonnes pratiques professionnelles**

#### **Projet d'analyse de données de bout en bout**

- Choix d'un jeu de données réel (finance, marketing, RH, industrie, santé...).
- Formulation des questions métiers et définition des indicateurs clés.
- Collecte, nettoyage, préparation et exploration des données.
- Construction de visualisations pertinentes et mise en place d'un modèle prédictif simple.
- Présentation des résultats sous forme de mini-rapport ou de soutenance.

#### **Bonnes pratiques professionnelles**

- Versioning de code avec Git (bases).
- Organisation des notebooks et des scripts pour faciliter la réutilisation.
- Éthique et confidentialité des données (RGPD, anonymisation, biais).

### **Préparation à la certification & examen**

#### **Structure de la certification**

- Présentation du format de l'examen "Analyste de données certifié Python – Niveau Associé".
- Domaines de compétences évalués : Python, pandas/NumPy, visualisation, statistiques, ML de base.
- Conseils pour bien gérer le temps et lire les énoncés.

## PROGRAMME DE LA FORMATION – DÉTAILLÉ

### **Révisions et mise en situation**

- Quiz de révision par module et correction commentée.
- Études de cas de type examen : analyse d'un dataset, interprétation de résultats, choix de visualisations.
- Session de questions/réponses avec le formateur.