

# Formation Python, Coder avec Pandas, NumPy & Scikit-learn

## Présentation

Ce cours intensif de 5 jours vous initie à l'utilisation de Python dans le contexte de la Data Science, en vous appuyant sur les bibliothèques incontournables :

- NumPy pour les calculs numériques,
- Pandas pour la manipulation de données tabulaires,
- Matplotlib/Seaborn pour la visualisation,
- Scikit-learn pour les modèles de machine learning.

À travers des cas concrets et des exercices pratiques, vous apprendrez à importer, transformer, visualiser et modéliser des jeux de données.

Durée : 35,00 heures (5 jours)

Tarif INTRA : Nous consulter

## Objectifs de la formation

À l'issue de la formation, les participants seront capables de :

- Utiliser NumPy pour les calculs numériques sur tableaux multidimensionnels
- Manipuler efficacement des données tabulaires avec Pandas
- Créer des visualisations claires avec Matplotlib et Seaborn
- Préparer les données pour la modélisation (feature engineering, normalisation)
- Appliquer des algorithmes de machine learning avec Scikit-learn
- Évaluer les performances d'un modèle prédictif

## Prérequis

- Bonnes bases en Python (variables, fonctions, boucles, listes)
- Connaissances de base en statistiques descriptives (moyenne, écart-type, corrélation...)
- Notions de manipulation de fichiers CSV ou Excel

## Public



- Analystes de données, data analysts
- Développeurs ou ingénieurs souhaitant se spécialiser en Data Science
- Responsables métier ou chefs de projet souhaitant monter en compétences sur la donnée
- Toute personne intéressée par l'analyse et la modélisation des données avec Python

## Programme de la formation

### Jour 1 : Fondamentaux et NumPy

- Introduction à Python pour la Data Science
- Jupyter Notebook et environnement Anaconda
- Introduction à NumPy
- Tableaux ndarray, indexation, slicing
- Fonctions mathématiques vectorisées
- Opérations sur matrices, statistiques de base
- Génération de données aléatoires (simulateurs, distributions)
- Manipulation d'axes, reshape, broadcast
- Introduction aux visualisations avec Matplotlib

### TP

- Simulation de données + opérations statistiques sur matrices

### Jour 2 : Manipulation de données avec Pandas

- Séries et DataFrames : création, import/export (CSV, Excel, JSON)
- Accès aux données (loc, iloc, at, iat)
- Nettoyage : gestion des valeurs manquantes, duplication, typage
- Filtres conditionnels, tri, regroupement (groupby)
- Agrégations et statistiques descriptives
- Merge, join, concaténation de DataFrames

### TP

- Nettoyage et analyse d'un jeu de données (ex : Airbnb, Titanic, données RH)

**Jour 3 : Visualisation de données**

- Visualisation avec Matplotlib : courbes, histogrammes, barres
- Introduction à Seaborn : styles, palettes, intégration avec Pandas
- Visualisation statistique : boxplots, heatmaps, pairplots
- Analyse exploratoire de données (EDA)
- Détection d'anomalies visuelles, corrélations, distribution
- Graphiques croisés dynamiques

**TP**

- Exploration graphique d'un dataset complet (ex : santé, marketing, transport)

**Jour 4 : Introduction au Machine Learning avec Scikit-learn**

- Introduction au Machine Learning : supervision vs non-supervision
- Séparation train/test, train\_test\_split, pipelines
- Modèles supervisés : régression linéaire, k-NN, arbre de décision
- Évaluation de modèle : précision, rappel, F1-score, courbe ROC
- Méthodes de validation croisée (cross\_val\_score)
- Feature scaling, normalisation

**TP**

- Prédiction d'un score (ex : prix d'un logement, revenu, classification binaire)

**Jour 5 : Atelier projet & approfondissements**

- Feature engineering : création de variables, encodage
- Algorithmes avancés (Random Forest, SVM – introduction)
- GridSearch et tuning des hyperparamètres

**Organisation**

**Formateur**

Les formateurs de Docaposte Institute sont des experts de leur domaine, disposant d'une expérience terrain qu'ils enrichissent continuellement. Leurs connaissances techniques et pédagogiques sont rigoureusement validées en amont par nos référents internes.

Riches de leur expérience sur le sujet, ils sauront accompagner vos collaborateurs dans leur montée en compétence.

**Moyens pédagogiques et techniques**

- Alternance d'apports théoriques et de mises en pratique contextualisées.
- Progression pédagogique structurée, de la compréhension à la mise en œuvre opérationnelle.
- Pédagogie immersive favorisant une mise en application immédiate des compétences.
- Formation interactive et participative s'appuyant sur des ateliers pratiques, études de cas et échanges collectifs (+ 60 % du temps de formation.)
- Séquences pédagogiques rythmées et évaluations formatives régulières pour renforcer l'ancrage des apprentissages.
- Supports de cours, ressources documentaires et références mis à disposition par le formateur

**Dispositif de suivi de l'exécution et de l'évaluation des résultats de la formation****En amont de la formation** *(évaluation en ligne dans l'extranet apprenant)*

- Recueil des besoins et des attentes des participants.
- Auto-positionnement permettant d'évaluer le niveau initial des apprenants.

**Tout au long de la formation**

- Évaluations continues réalisées par le formateur à travers des questions orales, exercices, QCM, études de cas et mises en situation pratiques.

**À l'issue de la formation** *(évaluations en ligne dans l'extranet apprenant)*

- Auto-positionnement afin de mesurer l'acquisition des compétences visées.
- Validation de l'acquisition des compétences par le formateur
- Questionnaire de satisfaction « à chaud » à l'issue de la formation

- Questionnaire de satisfaction "à froid", 45 jours après la formation afin d'évaluer l'ancrage des acquis et leur mise en application en situation professionnelle.

### **Accessibilité**

Nos formations peuvent être adaptées à certaines conditions de handicap. Nous contacter pour toute information et demande spécifique.