L'AKADÉMIE DU DIGITAL · TRAINING PROGRAMME Qualiopi certification being prepared

# Data Scientist

**FOR-DEV-M5-A · Development, DevOps & Data**

De la donnée au modèle : analyser, modéliser et prédire avec la science des données

DURATION
35 heures — 5 jours

PRICE
2?450 € HT

AUDIENCE
Schools & training centres

ACCESS
Access within 2 working days after validation of needs and prerequisites.

Tarif établi sur la base de 35h à 70 € HT/h. Parcours dédié aux écoles et centres de formation, établi sur devis ; tarif indicatif et négociable selon le volume et le contexte.

Le module alterne séquences théoriques et ateliers pratiques tout au long de chaque journée. Il couvre l'ensemble du cycle d'un projet data : manipulation et nettoyage des données, analyse exploratoire, statistiques, machine learning supervisé et non supervisé, évaluation et restitution. Le fil rouge est Python/scikit-learn, complété par un panorama de l'écosystème data.

## Learning objectives

- Comprendre le rôle du data scientist et le cycle de vie d'un projet data
- Manipuler, nettoyer et explorer des données avec Python (pandas, NumPy)
- Mener une analyse exploratoire et produire des visualisations pertinentes
- Comprendre les fondamentaux statistiques utiles à la modélisation
- Construire, entraîner et évaluer des modèles de machine learning (supervisé et non supervisé)
- Restituer et interpréter les résultats, en connaissant les limites, les biais et l'écosystème d'outils

## Target audience

- Étudiants et apprenants des écoles et centres de formation partenaires
- Profils techniques ou analytiques souhaitant s'orienter vers la data science
- Développeurs, analystes ou statisticiens en montée en compétence

## Prerequisites

Des bases de programmation sont requises (Python de préférence — module « Fondamentaux » recommandé) ainsi qu'une aisance avec la logique et les chiffres. Des notions de mathématiques et de statistiques facilitent la progression. Ces éléments sont validés par un formulaire en ligne et / ou un entretien préalable.

- Maîtriser les bases de la programmation (Python de préférence).

- Être à l'aise avec la logique, les chiffres et un raisonnement analytique.
- Avoir des notions de mathématiques / statistiques (un atout).
- Disposer d'une autonomie de travail et d'une curiosité pour la donnée.

## Programme

### Jour 1 — Le métier & manipulation des données

| | | | |
|---|---|---|---|
| Métier & écosystème data | 1h15 | Rôle du data scientist, cycle de vie d'un projet ; panorama des langages (Python, R, SQL) et environnements (Jupyter, Colab, Databricks).<br>Exposé interactif, panorama des outils, exemples de projets. | Questions orales ; QCM sur l'écosystème. |
| Prise en main de l'environnement | 2h15 | Découvrir l'environnement et charger des jeux de données.<br>TP guidé (Jupyter), import de données, prise en main. | Chargement de données validé. |
| Manipulation avec pandas | 1h00 | Structures de données, indexation, transformation.<br>Exposé, démonstration, exemples pandas/NumPy. | QCM sur pandas. |
| Transformer un jeu de données | 2h30 | Manipuler et transformer un jeu de données réel.<br>TP, jeu de données réel, opérations progressives. | Manipulations validées. |

### Jour 2 — Nettoyage & analyse exploratoire (EDA)

| | | | |
|---|---|---|---|
| Qualité & préparation des données | 1h15 | Valeurs manquantes, outliers, feature engineering.<br>Exposé, exemples de données sales, méthodes de nettoyage. | Questions orales ; QCM. |
| Nettoyer un jeu de données | 2h15 | Traiter manquants, outliers et incohérences.<br>TP guidé, jeu de données bruité, accompagnement. | Jeu de données nettoyé validé. |
| EDA & visualisation | 1h00 | Analyse exploratoire ; matplotlib, seaborn, ouverture Plotly et outils BI (Power BI, Tableau).<br>Exposé, exemples de visualisations, panorama des outils de viz. | QCM sur l'EDA et la visualisation. |
| Mener une EDA | 2h30 | Explorer et visualiser un jeu de données.<br>TP, production de graphiques, interprétation. | EDA et visualisations validées. |

### Jour 3 — Statistiques & fondamentaux du ML

| | | | |
|---|---|---|---|
| Statistiques pour la data science | 1h15 | Statistiques descriptives et inférentielles utiles.<br>Exposé, exemples chiffrés, interprétation. | Questions orales ; QCM. |
| Analyse statistique | 2h00 | Analyser statistiquement un jeu de données.<br>TP guidé, calculs et tests, interprétation. | Analyse statistique validée. |
| Principes du machine learning | 1h15 | Supervisé/non supervisé, train/test, surapprentissage.<br>Exposé, schémas, exemples, panorama scikit-learn / XGBoost / LightGBM. | QCM sur les principes ML. |
| Premier modèle | 2h00 | Construire un premier modèle simple (régression ou classification).<br>TP guidé scikit-learn, jeu de données préparé. | Premier modèle validé. |

### Jour 4 — Modélisation supervisée & évaluation

| | | | |
|---|---|---|---|
| Algorithmes supervisés | 1h15 | Régression, arbres, k-NN… ; pipeline de modélisation.<br>Exposé, comparaison d'algorithmes, schéma de pipeline. | Questions orales ; QCM. |
| Entraîner des modèles | 2h15 | Entraîner plusieurs modèles supervisés.<br>TP guidé scikit-learn, comparaison de modèles. | Modèles entraînés validés. |
| Évaluation & optimisation | 1h00 | Métriques, validation croisée, hyperparamètres.<br>Exposé, matrices de confusion, courbes, exemples. | QCM sur l'évaluation. |
| Évaluer & optimiser | 2h30 | Évaluer et optimiser un modèle.<br>TP, validation croisée, réglage des hyperparamètres. | Modèle évalué et optimisé validé. |

### Jour 5 — Non supervisé, interprétation & projet

| | | | |
|---|---|---|---|
| Non supervisé & éthique | 1h00 | Clustering, réduction de dimension ; biais, limites, éthique des modèles.<br>Exposé, exemples de clustering, cas de biais. | QCM sur le non supervisé et l'éthique. |
| Clustering | 2h00 | Appliquer un clustering et interpréter les résultats.<br>TP guidé, jeu de données, interprétation. | Clustering interprété validé. |
| Restitution & storytelling data | 1h00 | Restituer et communiquer les résultats aux décideurs.<br>Exposé, principes de data storytelling, exemples. | QCM sur la restitution. |
| Mini-projet data & restitution | 2h30 | Mener un mini-projet data de bout en bout et le présenter.<br>TP projet, présentations, revue collective, QCM final, enquête de satisfaction. | Évaluation sommative : projet data + QCM final. |

## Teaching methods and resources

In-person training at our premises or yours, in small groups. A mix of theory, hands-on workshops and practical scenarios.

**Required equipment :**

16 Go de Ram recommandés

- Les stagiaires doivent disposer d'un PC avec le Wifi et 16 Go de RAM recommandés.
- Effectif adapté au format école / centre de formation.
- Parcours dédié aux écoles et centres de formation, établi sur devis.

## Assessment methods

Learning outcomes are assessed during and at the end of the training; an attendance certificate is given to every learner.

## Accessibility

Training accessible to people with disabilities. Contact us before registering so we can study the adjustments needed.

## Funding

Quote or training agreement on request. Qualiopi certification is being prepared: eligibility for OPCO funding will open once it is obtained.

Certification audit planned for T1 2027. No OPCO eligibility before it is obtained.