L'AKADÉMIE DU DIGITAL · PROGRAMME DE FORMATION Certification Qualiopi en cours de préparation

Data Scientist

FOR-DEV-M5-A · Développement, DevOps & Data

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

DURÉE
35 heures — 5 jours

TARIF
2?450 € HT

PUBLIC
Écoles & centres de formation

DÉLAI D'ACCÈS
Accès sous 2 jours ouvrés après validation du besoin et des prérequis.

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.

Objectifs pédagogiques

- 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

Public concerné

- 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

Prérequis

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). 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. TP guidé (Jupyter), import de données, prise en main.	Chargement de données validé.
Manipulation avec pandas	1h00	Structures de données, indexation, transformation. Exposé, démonstration, exemples pandas/NumPy.	QCM sur pandas.
Transformer un jeu de données	2h30	Manipuler et transformer un jeu de données réel. 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. Exposé, exemples de données sales, méthodes de nettoyage.	Questions orales ; QCM.
Nettoyer un jeu de données	2h15	Traiter manquants, outliers et incohérences. 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). 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. 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. Exposé, exemples chiffrés, interprétation.	Questions orales ; QCM.
Analyse statistique	2h00	Analyser statistiquement un jeu de données. TP guidé, calculs et tests, interprétation.	Analyse statistique validée.
Principes du machine learning	1h15	Supervisé/non supervisé, train/test, surapprentissage. Exposé, schémas, exemples, panorama scikit-learn / XGBoost / LightGBM.	QCM sur les principes ML.
Premier modèle	2h00	Construire un premier modèle simple (régression ou classification). 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. Exposé, comparaison d'algorithmes, schéma de pipeline.	Questions orales ; QCM.
Entraîner des modèles	2h15	Entraîner plusieurs modèles supervisés. TP guidé scikit-learn, comparaison de modèles.	Modèles entraînés validés.
Évaluation & optimisation	1h00	Métriques, validation croisée, hyperparamètres. Exposé, matrices de confusion, courbes, exemples.	QCM sur l'évaluation.
Évaluer & optimiser	2h30	Évaluer et optimiser un modèle. 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. Exposé, exemples de clustering, cas de biais.	QCM sur le non supervisé et l'éthique.
Clustering	2h00	Appliquer un clustering et interpréter les résultats. TP guidé, jeu de données, interprétation.	Clustering interprété validé.
Restitution & storytelling data	1h00	Restituer et communiquer les résultats aux décideurs. 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. TP projet, présentations, revue collective, QCM final, enquête de satisfaction.	Évaluation sommative : projet data + QCM final.

Modalités et moyens pédagogiques

Formation en présentiel dans nos locaux ou dans vos locaux, en groupe restreint. Alternance d’apports théoriques, d’ateliers pratiques et de mises en situation.

Matériel requis :

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.

Modalités d’évaluation

Évaluation des acquis en cours et en fin de formation, attestation de fin de formation remise à chaque stagiaire.

Accessibilité

Formation accessible aux personnes en situation de handicap. Contactez-nous avant l’inscription pour étudier les aménagements nécessaires.

Financement

Devis ou convention de formation sur demande. La certification Qualiopi est en cours de préparation : l’éligibilité aux financements OPCO sera ouverte après son obtention.

Audit de certification prévu au T1 2027. Aucune éligibilité OPCO avant l’obtention.