

Les Essentiels du Data Scientist 100% Digital

Cette formation 100% en ligne vous formera au cœur du **métier de Data Scientist**. De la découverte du langage de programmation Python aux bonnes pratiques de DevOps. Vous pourrez suivre la formation à votre rythme, dans lequel nous recommandons qu'il nécessite un investissement de **130 heures sur 12 semaines**.

Cette formation est centrée sur le développement de compétences pratiques avec plus de 40 Notebooks en français, commentés avec des exercices et leurs corrections. avec **1 projet tutoré** à réaliser pour lesquels vous serez accompagné individuellement par notre équipe pédagogique composée de Seniors Data Scientists en poste en entreprise.

La formation se déroule **en distanciel** sous un **format asynchrone** où les vidéos explicatives, les quiz et les exercices d'applications sont au format digital.

Objectifs

Acquérir les **connaissances théoriques** et les **compétences pratiques** nécessaires pour exercer le métier de Data Scientist en entreprise. Ces compétences permettront de **modéliser des données structurées** sans biais et en utilisant le langage de programmation **Python**.

Compétences visées

- Maitriser le langage de programmation Python
- Manipuler et modéliser des données tabulaires
- Acquérir les connaissances théoriques clés sur les principaux algorithmes de Machine Learning supervisés & non-supervisés
- Implémenter en Python les principaux algorithmes de Machine Learning supervisés & non-supervisés
- Mettre en pratique les bonnes pratiques de développement informatique au sein de projet de Data Science

Public

Personnes en reconversion professionnelle, en recherche d'emploi, étudiants ou salariés ayant la volonté d'exercer un métier dans le domaine de la Data, en tant que Data Scientist.

Admission

Cette formation n'a pas de processus d'admission spécifique, ni de prérequis. N'importe qui souhaitant se former dès aujourd'hui en a la possibilité. Si vous avez les moindres questions, notre équipe pédagogique est disponible par téléphone.

DATAGONG - SAS au capital de 1000€

Siège Social : 17 rue Paul Verlaine, 94410 Saint-Maurice

SIRET : 913 142 055 00015 - APE : 85.59A

Enregistré auprès de la Direction Régionale et Interdépartementale de l'Économie, de l'Emploi, du Travail et des Solidarités (DRIEETS) d'Île-de-France sous le numéro 11941135494

DATAGONG est un Organisme de Formation Certifié QUALIOP1 au titre des actions de formation

Organisation

- Prochain lancement : **Dès maintenant**
- Rythme : **Avancez à votre rythme**. Les modules se débloquent au fur et à mesure de votre avancée. Néanmoins, nous recommandons **3 mois de formation (16 semaines)** réalisés en distanciel au travers de notre plateforme d'apprentissage
- Évaluation : Testez-vous tout au long de la formation au travers de nos **QUIZ**. Vous aurez la possibilité (non obligatoire) de réaliser 1 projet avec soutenance (150€ supplémentaires)
- Accompagnement : **Disponibilité quotidienne** de l'équipe pédagogique composée de Data Scientist
- Heures d'enseignement : **130 heures**

Programme

- Introduction au langage de programmation Python : 20 Heures
- Data Management en Python : 25 Heures
- Machine Learning en Python : 25 Heures
- Machine Learning Avancé en Python : 25 Heures
- Scoring en Python : 15 Heures
- Bonnes pratiques DevOps pour la Data Science : 20 Heures

Tarifs

499€ TTC

(Possibilité de financement par votre entreprise)

Contact

Email : contact@datagong.io

Tél : 06 62 06 05 25

Adresse : SAS DATAGONG, 17 rue Paul Verlaine, 94410 Saint-Maurice

Site internet : www.datagong.io

DATAGONG - SAS au capital de 1000€

Siège Social : 17 rue Paul Verlaine, 94410 Saint-Maurice

SIRET : 913 142 055 00015 - APE : 85.59A

Enregistré auprès de la Direction Régionale et Interdépartementale de l'Économie, de l'Emploi, du Travail et des Solidarités (DRIEETS) d'Île-de-France sous le numéro 11941135494

DATAGONG est un Organisme de Formation Certifié QUALIOP1 au titre des actions de formation

Détail du contenu et du créateur du cours

- **Introduction au langage de programmation Python** (bibliothèque NumPy) :
 - Les variables
 - Les contrôles logiques
 - Les listes
 - Les dictionnaires
 - Les boucles
 - Les fonctions
 - Les NumPy Arrays
 - Les fonctions scalaires et matricielles
 - Le broadcasting

Ce cours a été construit par **Paul LEYDIER** qui est **Lead Développeur Python** dans une fintech (Compass Financial Technologies) basée à Paris et à New-York. Paul est diplômé du master de Data Science de l'Université Paris 1 Panthéon-Sorbonne.

- **Data Management en Python** (bibliothèque pandas) :
 - Importer et explorer des DataFrames
 - Sélectionner et filtrer des données
 - Nettoyage des données pour la modélisation

Ce cours a été construit par **Paul THAVENOT** et **Guillaume CLÉMENT** qui sont **Lead Data Scientist** chez Orange et BNP Paribas et qui enseignent au Master de Data Science de l'Université Paris 1 Panthéon-Sorbonne. Paul et Guillaume sont diplômés du master de Data Science de l'Université Paris 1 Panthéon-Sorbonne.

- **Machine Learning en Python** (bibliothèque scikit-learn, NumPy) :
 - Apprentissage supervisé vs non-supervisé
 - Fonction de perte et optimisation des modèles
 - Compromis biais – variance
 - Validation-croisée
 - Apprentissage supervisé : Modèles de régression (régression linéaire, régression Ridge, régression Lasso, régression elastic-net)
 - Apprentissage supervisé : Modèles de classification (régression logistique)
 - Apprentissage non-supervisé : Réduction de dimensions (Analyse en Composantes Principales)
 - Apprentissage non-supervisé : Clustering (k-means, Classification Ascendante Hiérarchique)

Ce cours a été construit par **Amed COULIBALY** qui est **Lead Data Scientist** chez METRO et **Kaggle GrandMaster** (champion international en modélisation prédictive) et enseignant au master de Data Science de l'Université Paris 1 Panthéon-Sorbonne. Amed est diplômé du master de Data Science de l'Université Bretagne Sud.

- **Machine Learning Avancé en Python** (bibliothèques scikit-learn, xgboost, lightgbm) :
 - Classification binaire et multi-classes
 - Métriques d'évaluation
 - Utilisation des pipelines

DATAGONG - SAS au capital de 1000€

Siège Social : 17 rue Paul Verlaine, 94410 Saint-Maurice

SIRET : 913 142 055 00015 - APE : 85.59A

Enregistré auprès de la Direction Régionale et Interdépartementale de l'Économie, de l'Emploi, du Travail et des Solidarités (DRIEETS) d'Île-de-France sous le numéro 11941135494

DATAGONG est un Organisme de Formation Certifié QUALIOP1 au titre des actions de formation

- Support-Vector Machine (classification & régression)
- Arbre de décision (classification & régression)
- Modèles ensemblistes pour la classification et la régression (Bagging, Random Forest, AdaBoost, Gradient Boosting, xgboost, lightGBM)
- Méthodes pour limiter le sur-apprentissage et le sous-apprentissage
- Optimisation des hyper-paramètres

Ce cours a été construit par **Romain AVOUAC** qui est **administrateur de l'INSEE** en Data Science et enseignant à l'ENSAE Paris-Tech.

- **Scoring en Python** (bibliothèques scipy, scikit-learn, Boruta, lime, shap) :
 - Sélection de variables par des méthodes univariées
 - Sélection de variables par des méthodes multivariées (algorithme boruta)
 - Traitement des données déséquilibrées
 - Courbe de Lift
 - Calibration des scores (algorithme de Platt scaling, régression isotonique)
 - Interprétation globale et locale des modèles (méthodes Permutation feature importance & Partial dépendance plot, valeurs de Shapley, LIME)
 - Ciblage et monitoring

Ce cours a été construit par **Jules BONOGO** qui est **Lead Data Scientist** du centre de scoring de Orange. Jules est diplômé du master de l'ENSAE Paris-Tech Dakar et de l'ENSAI.

- **Bonnes pratiques DevOps pour la Data Science** :
 - Qualité du code avec pep8 et pep27
 - Architecture des projets avec Cookiecutter
 - Versionning et travail collaboratif avec GitHub
 - Portabilité d'une application avec les machines virtuelles et Docker
 - Introduction à la mise en production avec l'approche CI/CD

Ce cours a été construit par **Romain AVOUAC** qui est **administrateur de l'INSEE** en Data Science et enseignant à l'ENSAE Paris-Tech. Romain est diplômé de l'École Normale Supérieure de Lyon, de l'ENSAE Paris-Tech et de l'Université Paris 1 Panthéon-Sorbonne.

Évaluation

Comme précisé au chapitre « Organisation », **l'évaluation** se déroule de la manière suivante : Vous pourrez vous tester tout au long de la formation au travers de nos **QUIZ**. Vous obtiendrez à la fin de la formation vous obtiendrez une **attestation de suivi**.

IMPORTANT : Si vous souhaitez obtenir une **attestation de réussite**, cela passera par la réalisation d'un **projet avec soutenance**. Vous aurez 4 semaines pour le réaliser et il se finalisera par le passage d'une soutenance durant 30 minutes devant un Data Scientist.

Si vous souhaitez souscrire l'option soutenance (option non obligatoire), vous recevrez automatiquement à l'issue de la formation un email vous proposant de réaliser une soutenance pour un montant de **150€ TTC**. Cet email comportera l'ensemble des détails relatifs au projet. Si vous souscrivez à cette option, un nouvel espace se débloquera dans votre

DATAGONG - SAS au capital de 1000€

Siège Social : 17 rue Paul Verlaine, 94410 Saint-Maurice

SIRET : 913 142 055 00015 - APE : 85.59A

Enregistré auprès de la Direction Régionale et Interdépartementale de l'Économie, de l'Emploi, du Travail et des Solidarités (DRIEETS) d'Île-de-France sous le numéro 11941135494

DATAGONG est un Organisme de Formation Certifié QUALIOP1 au titre des actions de formation

espace sur la plateforme avec les instructions détaillées et les données. A l'issue des 4 semaines de réalisation du projet, nous déterminerons ensemble une date de soutenance.

Dans le cas où la soutenance n'est pas validée, vous aurez la possibilité de la repasser pour un montant de **50€ TTC**.

DATAGONG - SAS au capital de 1000€

Siège Social : 17 rue Paul Verlaine, 94410 Saint-Maurice

SIRET : 913 142 055 00015 - APE : 85.59A

Enregistré auprès de la Direction Régionale et Interdépartementale de l'Économie, de l'Emploi, du Travail et des Solidarités (DRIEETS) d'Île-de-France sous le numéro 11941135494

DATAGONG est un Organisme de Formation Certifié QUALIOP1 au titre des actions de formation