

• **DURÉE**

3 jours / 21h

APPRECIATION

Évaluation qualitative de fin de stage

MODALITÉS

ET MOYENS

PÉDAGOGIQUES

Démonstrations, cas pratiques, synthèse et évaluation des acquis

► **Objectifs**

pédagogiques

- Savoir modéliser un problème de Data Science en Python
- Découvrir les principales bibliothèques de calcul numérique dont Numpy, Pandas et Matplotlib

► **Public concerné**

Développeurs, Data Analystes, Data Scientists, et toute personne souhaitant acquérir des connaissances dans le domaine de la Data Science et Python

► **Prérequis**

Une première expérience en programmation Python. Avoir des connaissances de base en mathématiques et statistiques

Data Science PYTHON POUR LA DATA SCIENCE

2 030 € HT

Inter-Entreprises
Intra-entreprises sur devis

La Data Science nécessite des compétences spécifiques afin de prévoir le comportement des clients ou d'anticiper des tendances.

Cette formation vous permettra de découvrir la Data Science via Python, les différentes étapes qui interviennent lors de l'élaboration d'un modèle statistique, ainsi que l'analyse textuelle et la Data Visualisation.

Introduction à la Data Science

- Qu'est-ce que la Data Science ?
- Définition
- Domaines d'application de la Data Science
- Outils et algorithmes pour la Data Science
- De l'analyse statistique au Machine Learning
- Enjeux de la Data Science

Le langage Python

- Qu'est-ce que Python ?
- Shell, IDE, NoteBook
- Les bases de Python
- Structures de données
- Modules & fonctions
- Gestion des fichiers

Collecte et préparation des données avec Numpy et Pandas

- Importation et stockage de données
- Nettoyage et filtrage des données (valeurs manquantes, erronées...)
- Transformation des données
- Jointure & combinaison
- Agrégation & Grouping
- Exercices

Data Analysis – Modeling avec Scikit-Learn

- Introduction à Scikit-Learn
- Étapes d'élaboration d'un modèle
- Use case
- Apprentissage supervisé vs apprentissage non supervisé
- Evaluation du modèle
- Exercices

Data Visualisation avec Matplotlib

- Principe de la Data Visualisation
- Matplotlib : Scatter Plot, Box Plot, histogrammes
- Visualisation interactive
- Exercices

Introduction à l'utilisation de Spark avec Python

- Présentation de Spark
- Présentation de la librairie PySpark
- Les différents types d'analyses avec PySpark (SQL, Streaming, Machine Learning, Graphix)

Travaux pratiques

Alternance d'apports théoriques, d'exercices pratiques et de mise en situation sous forme de travaux pratiques, permettant de tester les différentes notions abordées.

