

R POUR LA DATA SCIENCE

Explorer et analyser les données passées et présentes pour prédire des données futures, tel est l'enjeu de la data science. Grâce à sa large variété d'outils statistiques et graphiques, R est un langage de référence pour les data scientists. Maîtriser la programmation en R fera de vous un acteur complet dans la réalisation de projets de data science.



Durée

- 2 jours



Tarif

- Inter : 950€ HT / participant (6 participants maximum)
- Intra : nous consulter



Lieu

- Locaux de Dcid Consulting (Orléans)
- Locaux de votre entreprise (Région Centre et Paris)



Public

Tout public ayant déjà quelques notions en programmation et/ou analyse de données

- Analyse de données ★★★★★
- Statistique ★★★★★
- Programmation ★★★★★
- Base de données ★★★★★
- Curiosité ★★★★★



Renseignements

- contact@dcid-consulting.fr
- 02 38 25 33 90



Prérequis

- Ordinateur individuel
- Connexion internet



Ressources utiles

- R, le site officiel
- Le big data.fr
- La fabrique du net.fr



Objectifs pédagogiques

- Maîtriser R, son fonctionnement et son écosystème
- Maîtriser le processus complet d'un projet de data science, de la collecte des données jusqu'à la communication des insights
- Analyser et exploiter des données structurées et non structurées avec R
- Elaborer des programmes d'analyses avec R



Programme

- Découvrir la science des données, en comprendre ses enjeux
- Découvrir R, son environnement et son écosystème
Les librairies, la programmation, les différents objets dans R
Les opérateurs, boucles, conditions et fonctions
- Importer, manipuler et exporter des données
Chargement de données de sources variées (texte, csv, Excel, SQL, web, etc.)
Préparation, extraction, filtrage et jointure du jeu de données
- Analyser et exploiter des données multivariées
Régression linéaire et régression multiple
Analyses factorielles et classification
- Analyser et exploiter des données textuelles
Import et manipulation de données textuelles
Les principaux concepts : tokenization, lemmatisation, stop-words, etc.
- Modéliser et prédire l'information
Apprentissage supervisé et non supervisé
Introduction au machine learning
- Visualiser et communiquer ses résultats
Les principes de la dataviz, les différents type de graphique et leurs usages
La librairie ggplot2 et les librairies interactives
- Et après ? Perspectives et cas d'usage



Méthodes et moyens pédagogiques

- Alternance d'exposés et d'applications pratiques
- Possibilité de travailler sur les jeux de données des participants
- Codes source et supports de formation remis aux participants
- Formation dispensée par un data scientist expérimenté (pratique de R depuis 7 ans minimum)