

FORMATION R POUR LA DATA SCIENCE

Type de formation : R101

DESCRIPTION

Trois jours pour vous familiariser avec le langage R et toutes les subtilités de son utilisation pour la data science. Vous sortirez de cette formation avec une maîtrise de l'utilisation et des principes de base du logiciel R pour la data science.

OBJECTIFS

- Maîtrise de l'environnement de travail de R avec Rstudio
- Manipulation et gestion de données avec R et les packages de data manipulation les plus récents (dplyr, data.table...)
- Maîtrise de l'application de méthodes de data science
- Connaître les bases de la programmation avec R

DUREE

3 jours

PUBLIC ET PREREQUIS

Aucun prérequis sur l'utilisation de R n'est nécessaire. Quelques bases en traitement des données.

MOYENS PEDAGOGIQUES

Alternance d'exposés et d'applications pratiques avec des exercices sur des données.

PROGRAMME DETAILLE

- **Le langage R et son environnement :**
 - Définition de l'environnement
 - Présentation de la notion de packages
 - Présentation de RStudio
 - Les possibilités de R pour la création de documents (scripts, Rmarkdown, Rnotebook...)
 - Présentation des packages importants pour la manipulation de données
- **Les structures et les fonctionnalités de base de R :**
 - L'importation des données (.txt, .csv, .xls, .sav...)
 - Les fonctionnalités de base pour la gestion des données
- **Utilisation avancée de la gestion des données avec R :**
 - La grammaire de dplyr
- **Statistiques descriptives avec R**
 - Les fonctionnalités pour la description statistique des données
 - La gestion des graphiques (histogrammes, box-plot...)
- **Le data mining et le machine learning avec R**
 - L'analyse descriptive des données multivariées et la réduction de dimensions (ACP)
 - Présentation et utilisation du package FactoMineR
 - Apprentissage supervisé – cas pratiques avec R
 - La régression logistique

STAT4DECISION

37-39 avenue Ledru-Rollin – 75012 Paris – France

Tel. 01.72.25.40.82 – E-mail : info@stat4decision.com – www.stat4decision.com

SIRET : 810 489 856 00015 – Numéro de déclaration d'activité : 11 75 53522 75

stat4decision est validé dans le DATA-DOCK

- Les arbres de décisions
- Les forêts aléatoires et les GBM
- Présentation du package caret pour ajuster les hyper-paramètres
- L'apprentissage non supervisé – cas pratiques avec R (k-means...)
- **Quelques notions de programmation avec R**
 - Les standards de la programmation avec R
 - Construire une fonction R pour automatiser des analyses
 - Les structures de contrôle : if, for, while...