LANGAGE R INITIATION

Programme complet

1) Les fondamentaux de la programmation R pour la Data Science

- Découverte de R

1. Présentation de R
2. Installation et ressources
3. Le système de packages
4. IDE RStudio et création d'un premier script

- Les bases du langage R

1. Les variables et les types
2. Les formats : les vecteurs
3. Les formats : les matrices
4. Les formats : les DataFrame
5. Les formats : les listes
6. Les conditions (if ... else)
7. Les boucles
8. Les fonctions
9. Les fonctions de type apply

- R et la data

1. Récupérer les données d'un fichier au format CSV
2. Récupérer d'autres formats de données : Excel, SQL, API

- Analyses statistiques et premiers graphiques

1. Analyse de variables numériques
2. Analyse de variables catégorielles
3. Autres méthodes d'analyses statistiques

- Restitution de résultats : aperçu des possibilités

1. Les graphiques
2. Les cartes
3. La création de reporting avec le package rmarkdown
4. La création de tableaux de bord web avec le package shiny

2) Manipulez vos données avec les packages dplyr et tidyr

- Rappels sur l'accès avec R à différents formats de données

1. Accès à des données au format CSV
2. Accès à d'autres formats de données (Excel, bases de données SQL, API...)

- Manipulation à l'aide du package dplyr

1. Introduction au package "dplyr" : syntaxe et "pipe"
2. Manipulation d'observations (lignes)
3. Manipulation de variables (colonnes)
4. Création des tableaux de données résumées
5. Jointures de données

Objectifs pédagogiques

- Utiliser l'environnement R
- Manipuler des données et des objets
- Programmer avec R
- Appliquer des méthodes d'analyses statistiques courantes

Public concerné

Tout développeur en environnement statistique, personnes voulant découvrir le datamining et l'appliquer.

Pré-requis

Avoir des connaissances de base en statistiques (régression linéaire, échantillonnage) ainsi que des connaissances de base en programmation (variables, boucles, etc.).

Durée

20h00

Moyens Pédagogiques

Formation « MOOC », tutoring vidéo et mise en application

Livre broché 748 pages + support numérique en ligne + Vidéos

Le formateur travaille dans la data depuis 10 ans. En tant que Data Scientist, il conçoit dans le cloud des produits logiciels basés sur des analyses statistiques de données financières, spatiales et temporelles.

Lorsqu'il vous accompagne dans l'apprentissage du langage R, c'est toute son expérience qu'il partage : concevoir des indicateurs pour analyser ses données, prendre des décisions éclairées et quantifier leur impact

Evaluation du contrôle des connaissances

Le formateur évalue la progression pédagogique du participant tout au long de la formation au moyen de mises en situation et de travaux pratiques.

Mode de validation de la formation

Attestation de formation.

- Introduction au nettoyage de données avec le package tidyr

1. Remodeler des données
2. Compléter les valeurs manquantes

3) Prendre en main les statistiques

- Quelques rappels sur R et son environnement

1. Pourquoi R ?
2. Installation de l'environnement
3. Ressources liées à R

- Données au format CSV

1. Accéder à des données au format CSV
2. Mise en pratique - Data Visualisation : graphique en bâtons

- Données au format Excel

1. Accéder à des données au format Excel
2. Mise en pratique - Data Visualisation : nuage de points

- Données d'une base de données SQL

1. Accéder à des données provenant d'une base de données SQL
2. Mise en pratique - Data Visualisation : nuage de mots

- Données provenant d'une API

1. Accéder à des données provenant d'une API
2. Mise en pratique - Data Visualisation : graphique dynamique en bâtons pour le web

- Données provenant d'une base de données NoSQL MongoDB

1. Accéder à des données provenant d'une base de données NoSQL
2. Mise en pratique - Data Visualisation : tableau dynamique pour le web