Analyse et automatisation de données avec R

Référence : **BDDR001** Durée : **2 jours (14 heures)** Certification : **Aucune**

CONNAISSANCES PRÉALABLE

- Posséder des notions de base en statistiques
- Avoir une première pratique du langage R

PROFIL DES STAGIAIRES

- Data analysts, statisticiens

OBJECTIFS

- Manipuler et transformer des données
- Visualiser des données avec ggplot2
- Automatiser des analyses statistiques

CERTIFICATION PRÉPARÉE

- Aucune

MÉTHODES PÉDAGOGIQUES

- Mise à disposition d'un poste de travail par personne
- Remise d'une documentation pédagogique numérique pendant le stage
- La formation est constituée d'apports théoriques interactif, démonstrations en direct, travaux dirigés, ateliers collaboratif, mises en situation professionnelles, projet fil rouge, études de cas réels, ateliers de préparation et de qualité des données, ateliers de datavisualisation et de communication des résultats, ateliers d'automatisation des analyses
- Le suivi de cette formation donne lieu à la signature d'une feuille d'émargement par demi-journée de formation

FORMATEUR

- Consultant-formateur expert en analyse et automatisation des données avec R

MÉTHODES D'ÉVALUATION DES ACQUIS

- Validation des prérequis envoyée en amont de la formation. En l'absence de réponse d'un ou plusieurs participants, un temps sera consacré en ouverture de session pour prendre connaissance du positionnement de chaque stagiaire sur les objectifs pédagogiques évalués
- Validation progressive des compétences tout au long de la formation via exercices pratiques, quiz techniques, études de cas, analyses statistiques, créations de visualisations, automatisation de traitements et restitution des travaux réalisés dans le cadre du projet fil rouge
- Auto-évaluation des acquis par le stagiaire via un questionnaire
- Attestation des compétences acquises envoyée au stagiaire
- Attestation de fin de stage adressée avec la facture

ACCESSIBILITÉ DE LA FORMATION

- Si vous avez besoin d'un aménagement pour suivre votre formation dans de bonnes conditions, nous vous invitons à contacter notre référent handicap au 01 71 19 70 30 ou à l'adresse referent.handicap@edugroupe.com. Un échange en amont nous permettra d'étudier avec vous les adaptations possibles.

CONTENU DU COURS

1. JOUR 1 - Préparer, transformer et analyser les données avec R (7h)

2. Prendre en main l'environnement R et le Tidyverse (2h)

- Présentation de R et RStudio
- Organisation d'un projet d'analyse
- Packages et gestion des dépendances
- Introduction au Tidyverse
- Importation de données CSV, Excel et JSON
- Principes de reproductibilité
- 💡 *Mises en pratique : Installation et configuration de l'environnement ; Création d'un projet RStudio ; Importation et exploration d'un jeu de données*

3. Manipuler et préparer les données avec dplyr et tidyr (3h)

- Sélection et filtrage
- Tri et regroupement
- Création de variables calculées
- Agrégations
- Jointures
- Restructuration des données
- Gestion des valeurs manquantes
- Contrôle qualité des données
- 💡 *Mises en pratique : Nettoyage d'un dataset ; Fusion de plusieurs sources ; Construction d'une table analytique*

4. Réaliser des analyses statistiques avec R (2h)

- Statistiques descriptives
- Distribution des variables
- Corrélations
- Comparaison de populations
- Analyse exploratoire
- Premiers indicateurs métiers
- 💡 *Mises en pratique : Analyse statistique d'un jeu de données ; Interprétation des résultats ; Construction d'indicateurs de synthèse*

5. JOUR 2 - Visualiser et automatiser les analyses (7h)

6. Créer des visualisations avec ggplot2 (3h)

- Principes de datavisualisation
- Graphiques en barres
- Histogrammes
- Courbes
- Nuages de points
- Boxplots
- Personnalisation des graphiques
- Storytelling visuel
- 💡 *Mises en pratique : Création de visualisations analytiques ; Mise en forme de graphiques ; Construction d'un tableau de bord simple*

7. Automatiser les traitements avec R (2h)

- Fonctions personnalisées
- Scripts automatisés
- Paramétrage des traitements
- Gestion des erreurs
- Réutilisation du code
- Bonnes pratiques d'automatisation
- 💡 *Mises en pratique : Création de scripts réutilisables ; Automatisation d'une chaîne de traitement ; Paramétrage d'analyses*

8. Produire des rapports automatisés avec R Markdown (2h)

- Introduction à R Markdown
- Génération de rapports HTML et PDF
- Intégration des analyses et graphiques
- Paramétrisation des rapports
- Bonnes pratiques de documentation
- 💡 *Mises en pratique : Création d'un rapport automatisé ; Génération d'un livrable complet ; Restitution des résultats et débriefing collectif*