Institut Régional de Formation des Allocations Familiales
67 avenue Jean Jaurès - 75019 PARIS CEDEX 19 - Tél. : 01 71 13 36 18
Siret : 381 050 996 00127 – APE 8559 A – N° déclaration d'activité : 11 75 48596 75

Perspicaf PE-RINI : Initiation à la programmation avec R sous Databricks

Description :

La modernisation du Système d'informations décisionnelles (SID) de la Branche Famille nécessite le remplacement de l'outil actuel SEG et son langage de requêtes SAS, par de nouveaux outils de traitement des données, comme Power BI et la plateforme Azure Databricks. Celle-ci requière l'utilisation des langages SQL, Python et, si besoin, R pour manipuler les données de la Branche Famille. Cette action propose aux chargés d'études de les initier à l'utilisation du langage R sous Databricks pour extraire et manipuler les données du SID.

Objectifs :

- Extraire des données du SID via Sparklyr
- Manipuler les données et agréger les statistiques
- Réaliser des restitutions sous forme de tableaux et de listings

Programme :

- Environnement Databricks et bases du langage R
  - Databricks : quelques points de repère
  - Architecture : où sont les données ? qu'est-ce que le calcul distribué ?
  - Notion de package R, chargement
  - Objets, affectation, logique du code R
  - Connexion à un cluster avec {sparklyr} et charger des données
  - Import, export
- Manipulation de données avec {sparklyr}
  - Rôle du connecteur %>%
  - Sélections de colonnes
  - Filtres sur les lignes
  - Agrégations et statistiques descriptives : calculs d'indicateurs
  - Jointures
  - Création / modification de colonnes : manipulations de textes, de dates, de catégories, recodages
  - Tris
  - Doublons : identification, élimination
  - Transformation du résultat en data.frame dans R
- Exploitation des données : graphiques
  - Principe du package {ggplot2}
  - Diagrammes en bâtons

- Courbes et nuages de points
- Boîtes à moustaches
- Graphiques multiples

Méthode pédagogique :

- Apports théoriques
- Exemples concrets en Caf
- Exercices d'application

Modalités d'évaluation et de validation :

- **Evaluation de positionnement** : sous forme d'un questionnaire ou d'un tour de table avec le formateur pour valider les prérequis, pour évaluer les besoins individuels et pour déterminer le niveau de connaissances
- **Evaluation des acquis** : validation de la compréhension et de l'acquisition des connaissances sous forme de mises en situations, de réflexions collectives et d'utilisation d'outils de diagnostic
- **Evaluation à chaud :** à la fin de la formation, un bilan oral est effectué par le formateur et une évaluation écrite adressée aux stagiaires permettent d'apprécier la qualité de la prestation et de mesurer l'efficacité de l'action au regard des objectifs globaux
- **Evaluation à froid** : réalisée avec un outil interne Caf
- **Attestation de suivi** : Feuille de présence
- **Certificat de réalisation** mentionnant la nature, la durée de l'action est remis aux stagiaires à l'issue de la formation

Accessibilité :

Nous mettons tout en œuvre afin d'offrir aux personnes en situation de handicap des conditions optimales d'accès et d'apprentissage. N'hésitez pas à contacter Naima Ouari référent handicap naima.ouari@caf92.caf.fr - 01 87 02 85 25 / 06 09 28 97 89 directement pour lui signaler vos besoins spécifiques.

Public cible :
Les chargés d'études dont l'activité nécessite l'utilisation du langage R
Pré-requis :

- Utilisation des données du SID
- Utilisation de l'environnement Databricks

Programmation :
Nous contacter Planification sur le site irfaf.fr
Catégorie :
PERSPICAF

Coût forfait / stagiaire : 1100.00 €
Coût par forfait / groupe : 6000.00 €
Modalités animation :

- Présentiel

Formateur :
Effectif :
6 à 10 participants
Durée en jours :
3
Durée en heures :

Lieu : 21
Irfaf - 67 Avenue Jean Jaurès 75019 Paris