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

ASM : Analyse Statistique Multidimensionnelle avec Python dans Databricks

Description :

Le métier de chargés d'études consiste à manipuler des données brutes dans le but de réaliser des analyses sociologiques, statistiques et démographiques afin notamment d'apporter les éléments pour aider à la décision. Ces analyses requièrent des compétences à la fois théoriques, méthodologiques et pratiques en programmation.
En outre, les fonctionnalités de Databricks offrent de nouvelles perspectives d'analyse de données plus avancées (régression, modèles de prévision, machine Learning, analyse géospatiale, ...). Les chargés d'études doivent donc être formés sur les méthodes statistiques théoriques et pratiques afin de réaliser des ACM et AFC sous Databricks (utilisation du langage Python et des bibliothèques Pandas et matpotlib),

Objectifs professionnels

- Analyser les caractéristiques socio-économiques des allocataires à l'aide d'une Analyse des Correspondances Multiples (ACM), afin d'identifier des profils types et d'orienter les actions de communication ou d'accompagnement vers les publics les plus vulnérables.
- Mettre en œuvre des outils d'exploration et de visualisation des données (avec Pandas et Matplotlib) pour produire des tableaux de bord destinés à suivre l'évolution des prestations familiales par territoire ou par typologie de ménage.

Objectifs :

- utiliser les librairies Pandas et Matplotlib sous Python pour charger, filtrer, transformer et explorer un jeu de données, puis produire des visualisations pertinentes (histogrammes, boîtes à moustaches, nuages de points, courbes temporelles) permettant de mettre en évidence les principales tendances et anomalies.
- mettre en œuvre une Analyse Factorielle des Correspondances (AFC) et une Analyse des Correspondances Multiples (ACM) à l'aide du langage Python, d'en interpréter les résultats et de visualiser les dimensions principales pour faciliter la compréhension des relations entre modalités.

Programme :

- Rappel des fondamentaux statistiques
- Tests statistiques essentiels
- Analyse statistique multidimensionnelle
  - Analyse Factorielle des Correspondances (AFC)
  - Analyse de Correspondances Multiples (ACM)

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 Sandrine Decottignies référent handicap sandrine.decottignies@caf92.caf.fr - 01 87 02 85 05 directement pour lui signaler vos besoins spécifiques.

Public cible :
Chargés d'études, Data Analyste, Data Scientiste, Contrôleurs de gestion ayant des connaissances initiales en statistiques
Pré-requis :

- Maitriser le langage python dans Databricks,
- Maîtriser les méthodes d'échantillonnage aléatoire simple et stratifié et avoir mis en application la fiche Statistiques *n°09 - Echantillonage avec sample_spark*,
- Maitriser les concepts de statistique descriptive (variable quali/quanti, moyenne, quantiles, variance, écart-type...),
- Connaître les grands principes et méthodes de statistique décisionnelle (tests, estimations, régression) et d'analyse des données multidimensionnelles,
- Utiliser les données du Sid de la branche Famille.

Programmation :

Coût forfait / stagiaire : 1459.00 €
Coût par forfait / groupe : 7295.00 €
Modalités animation :

- Présentiel

Formateur :
Effectif :
5 à 8 participants
Durée en jours :
2,5
Durée en heures :
17h30

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

Nous contacter Planification sur le site irfaf.fr

Catégorie :

Formation PERSPICAF - Métiers de l'observation socio-économique et de l'appui au pilotage

Lieu :

IRFAF - 67 avenue Jean Jaurès 75019 Paris