Analyse de données avec Python – intermédiaire

Référence : **BDDPYT004** Durée : **3 jours (21 heures)** Certification : **Aucune**

CONNAISSANCES PRÉALABLE

- Avoir des connaissances de bases de Python et d'Excel

PROFIL DES STAGIAIRES

- Analystes, data scientists

OBJECTIFS

- Manipuler les données avec pandas
- Visualiser les données avec matplotlib et seaborn
- Nettoyer et préparer les datasets
- Réaliser des analyses statistiques de base

CERTIFICATION PRÉPARÉE

- Aucune

MÉTHODES PÉDAGOGIQUES

- Mise à disposition d'un poste de travail par stagiaire
- 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 storytelling analytique
- 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 de données et Python

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, analyses de jeux de données, création de visualisations, interprétation de résultats statistiques 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 et manipuler les données avec Python (7h)

2. Rappels Python et découverte de l'écosystème Data (2h)

- Rappels Python utiles à l'analyse de données
- Jupyter Notebook
- NumPy
- Pandas
- Matplotlib
- Seaborn
- Organisation d'un projet d'analyse
- Bonnes pratiques de manipulation des données
- 💡 *Mises en pratique : Configuration de l'environnement ; Manipulation de notebooks ; Exploration d'un premier dataset*

3. Manipuler les données avec pandas (3h)

- DataFrame et Series
- Import de fichiers CSV, Excel et JSON
- Sélection et filtrage
- Tri des données
- Création et modification de colonnes
- Agrégations
- GroupBy
- Fusion et jointure de datasets
- 💡 *Mises en pratique : Analyse exploratoire de données ; Création de tableaux de synthèse ; Fusion de plusieurs sources de données*

4. Nettoyer et préparer les données (2h)

- Valeurs manquantes
- Doublons
- Détection des incohérences
- Conversion de types
- Standardisation des formats
- Gestion des données aberrantes
- Préparation des données pour l'analyse statistique
- 💡 *Mises en pratique : Nettoyage complet d'un dataset ; Correction des anomalies ; Contrôle qualité des données*

5. JOUR 2 - Explorer et visualiser les données (7h)

6. Réaliser une analyse exploratoire des données (EDA) (3h)

- Statistiques descriptives
- Distribution des variables
- Corrélations
- Détection d'outliers
- Analyse univariée
- Analyse bivariée
- Premiers indicateurs métiers
- 💡 *Mises en pratique : Réalisation d'une analyse exploratoire complète ; Identification des tendances ; Interprétation des résultats*

7. Créer des visualisations avec Matplotlib (2h)

- Principes de datavisualisation
- Histogrammes
- Courbes
- Diagrammes en barres
- Graphiques de dispersion
- Personnalisation des graphiques
- Export des visualisations
- 💡 *Mises en pratique : Création de plusieurs visualisations ; Mise en forme des graphiques ; Construction d'un rapport graphique*

8. Réaliser des visualisations avancées avec Seaborn (2h)

- Introduction à Seaborn
- Heatmaps
- Pairplots
- Boxplots
- Violin Plots
- Matrices de corrélation
- Analyse visuelle multivariée
- 💡 *Mises en pratique : Création de visualisations avancées ; Analyse de corrélations ; Présentation graphique de résultats*

9. JOUR 3 - Réaliser des analyses statistiques et projet de synthèse (7h)

10. Réaliser des analyses statistiques avec Python (3h)

- Moyenne
- Médiane
- Variance
- Écart-type
- Quartiles
- Corrélations
- Analyse de distributions
- Introduction aux tests statistiques simples
- 💡 *Mises en pratique : Calcul d'indicateurs statistiques ; Analyse comparative de populations ; Interprétation des résultats*

11. Produire des rapports d'analyse reproductibles (2h)

- Organisation des notebooks
- Documentation des analyses
- Export des résultats
- Présentation des conclusions
- Bonnes pratiques de reproductibilité
- 💡 *Mises en pratique : Création d'un notebook d'analyse complet ; Structuration d'un rapport de résultats ; Préparation d'une restitution*

12. Projet fil rouge : analyse complète d'un jeu de données métier (2h)

- Compréhension du besoin
- Préparation des données
- Nettoyage
- Analyse exploratoire
- Visualisation
- Analyse statistique
- Synthèse des résultats
- Recommandations
- 💡 *Mises en pratique : Réalisation d'un projet complet d'analyse de données ; Production de visualisations et indicateurs ; Présentation des conclusions ; Débriefing collectif et retour d'expérience*