Analyse de données avec Python – avancé

Référence : **BDDPYT005** Durée : **3 jours (21 heures)** Certification : **Aucune**

CONNAISSANCES PRÉALABLE

- Avoir suivi la formation Analyse de données avec Python niveau intermédiaire ou posséder les connaissances et compétences équivalentes

PROFIL DES STAGIAIRES

- Data scientists, analystes expérimentés

OBJECTIFS

- Créer des visualisations avancées
- Utiliser des techniques statistiques avancées
- Manipuler de grands volumes de données
- Automatiser des analyses avec Python

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 d'analyse statistique avancées, ateliers d'automatisation et d'industrialisation 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 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, 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 - Manipulation avancée et traitement de grands volumes de données (7h)

2. Optimiser la manipulation de données avec pandas et NumPy (2h30)

- Rappels avancés sur les DataFrames
- Index multi-niveaux
- Reshape des données
- Pivot et Pivot Table
- GroupBy avancés
- Fenêtres glissantes
- Fonctions d'agrégation personnalisées
- Optimisation des traitements avec NumPy
- 💡 *Mises en pratique : Construction de tableaux analytiques complexes ; Agrégations avancées ; Optimisation de traitements existants*

3. Manipuler de grands volumes de données (2h)

- Optimisation mémoire
- Types de données pandas
- Lecture par chunks
- Traitements incrémentaux
- Formats Parquet et Feather
- Introduction à Dask
- Bonnes pratiques Big Data avec Python
- 💡 *Mises en pratique : Traitement de fichiers volumineux ; Optimisation de DataFrames ; Comparaison de performances*

4. Automatiser les traitements de données (2h30)

- Création de fonctions réutilisables
- Scripts automatisés
- Gestion des paramètres
- Modularisation du code
- Gestion des logs
- Génération automatique de rapports
- Introduction à l'orchestration des traitements
- 💡 *Mises en pratique : Création d'un pipeline automatisé ; Génération automatique de rapports d'analyse ; Mise en place d'un traitement réutilisable*

5. JOUR 2 - Statistiques avancées et exploration approfondie des données (7h)

6. Réaliser des analyses statistiques avancées (3h)

- Tests statistiques
- Comparaison de populations
- Tests de normalité
- Tests paramétriques et non paramétriques
- Intervalles de confiance
- Analyse de corrélations
- Introduction aux régressions statistiques
- 💡 *Mises en pratique : Réalisation de tests statistiques ; Analyse comparative de jeux de données ; Interprétation des résultats*

7. Analyse exploratoire avancée (EDA avancée) (2h°

- Analyse multivariée
- Détection avancée d'outliers
- Analyse de distributions
- Segmentation exploratoire
- Réduction de dimensions (introduction PCA)
- Identification de patterns
- 💡 *Mises en pratique : Analyse approfondie d'un dataset complexe ; Détection de comportements atypiques ; Analyse de structures de données*

8. Construire des indicateurs et tableaux d'analyse avancés (2h)

- KPI analytiques
- Construction d'indicateurs composites
- Analyse temporelle
- Analyse comparative
- Segmentation
- Synthèse décisionnelle
- 💡 *Mises en pratique : Création de tableaux d'indicateurs ; Analyse métier de données ; Construction de tableaux de synthèse*

9. JOUR 3 - Datavisualisation avancée et projet de synthèse (7h)

10. Réaliser des visualisations avancées avec Matplotlib et Seaborn (3h)

- Personnalisation avancée des graphiques
- Visualisations multivariées
- Heatmaps
- Pairplots
- Boxplots avancés
- Visualisations temporelles
- Analyse graphique de corrélations
- Storytelling visuel
- 💡 *Mises en pratique : Création d'un ensemble de visualisations avancées ; Analyse visuelle de phénomènes complexes ; Construction d'un rapport graphique*

11. Créer des rapports analytiques reproductibles (2h)

- Organisation avancée des notebooks
- Documentation des analyses
- Génération de rapports automatisés
- Export des résultats
- Bonnes pratiques de reproductibilité
- Versionnement des analyses
- 💡 *Mises en pratique : Création d'un notebook professionnel ; Génération automatisée d'un rapport ; Structuration d'une analyse reproductible*

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

- Analyse du besoin métier
- Préparation et optimisation des données
- Analyse exploratoire avancée
- Analyses statistiques
- Création de visualisations
- Construction d'indicateurs
- Automatisation des traitements
- Synthèse et recommandations
- Mises en pratique : Réalisation d'un projet complet d'analyse avancée ; Production d'un rapport d'analyse automatisé ; Présentation des conclusions et recommandations ; Débriefing collectif et retour d'expérience