

BASES DE DONNÉES ET SQL POUR LA SCIENCE DES DONNÉES AVEC PYTHON

Durée : 3 jours (21 heures)

CONNAISSANCES PREALABLES

- Connaissances générales en informatique
- Notions de programmation appréciables
- Connaissances élémentaires en Python recommandées
- Aucune expertise préalable en bases de données n'est requise.

PROFIL DES STAGIAIRES

- Data Analysts
- Data Scientists débutants
- Data Engineers juniors
- Développeurs Python
- Consultants Data et BI
- Ingénieurs souhaitant évoluer vers les métiers de la donnée
- Étudiants ou professionnels en reconversion vers la Data
- Toute personne souhaitant exploiter des bases de données avec SQL et Python.

OBJECTIFS

À l'issue de cette formation, les participants seront capables de :

- Comprendre le fonctionnement des bases de données relationnelles
- Modéliser simplement des données relationnelles
- Interroger une base de données avec SQL
- Réaliser des analyses de données à l'aide de requêtes SQL
- Connecter Python à une base de données relationnelle
- Manipuler les données avec Pandas
- Automatiser l'extraction et l'analyse des données
- Construire un flux simple d'analyse de données avec Python et SQL.

CERTIFICATION PREPAREE

Aucune

METHODES PEDAGOGIQUES

- Alternance d'apports théoriques et de travaux pratiques
- Exercices progressifs SQL
- Manipulations Python et Pandas
- Études de cas Data
- Cas fil rouge couvrant l'ensemble de la formation

- Analyse de jeux de données réels.

FORMATEUR

- Consultant expert Data Science, Data Engineering et bases de données, spécialisé dans les environnements Python, SQL et l'analyse de données.

METHODE D'EVALUATION DES ACQUIS

- Quiz de validation des connaissances
- Exercices SQL
- Travaux pratiques Python
- Études de cas
- Projet fil rouge
- Évaluation continue des acquis.

CONTENU DU COURS

Module 1 – Comprendre les bases de données relationnelles (2h)

Objectifs

- Comprendre le fonctionnement d'un SGBDR
- Maîtriser les concepts fondamentaux du modèle relationnel

Contenu

- Bases de données et systèmes d'information
- Tables, colonnes et enregistrements
- Clés primaires et étrangères
- Relations entre les données
- Présentation des principaux SGBDR :
 - PostgreSQL
 - MySQL
 - SQL Server
 - Oracle
- Cycle de vie des données

Mise en pratique

- Analyse d'un schéma relationnel
- Identification des relations entre tables
- Étude de cas métier

Module 2 – Interroger les données avec SQL (4h)

Objectifs

- Construire des requêtes SQL de consultation
- Extraire des informations pertinentes

Contenu

- Syntaxe SQL fondamentale
 - SELECT
 - WHERE
 - ORDER BY
 - DISTINCT
 - Alias
-

- Fonctions de calcul
- Expressions SQL
- Gestion des valeurs NULL

Mise en pratique

- Construction de requêtes simples
- Extraction ciblée de données
- Exercices progressifs d'analyse

Module 3 – Agrégation et analyse des données avec SQL (3h)**Objectifs**

- Produire des indicateurs et des synthèses
- Réaliser des analyses exploratoires

Contenu

- COUNT
- SUM
- AVG
- MIN
- MAX
- GROUP BY
- HAVING
- Calcul d'indicateurs
- Analyse statistique simple

Mise en pratique

- Création de tableaux de synthèse
- Calcul d'indicateurs métier
- Études de cas Data

Module 4 – Jointures et requêtes avancées (3h)**Objectifs**

- Exploiter des bases relationnelles multi-tables
- Réaliser des analyses complexes

Contenu

- INNER JOIN
- LEFT JOIN
- RIGHT JOIN
- FULL JOIN
- Sous-requêtes
- Requêtes imbriquées
- Bonnes pratiques SQL

Mise en pratique

- Construction de requêtes multi-tables
- Analyse de jeux de données complexes
- Exercices avancés

Module 5 – Connecter Python aux bases de données (2h)**Objectifs**

- Lire des données SQL depuis Python
- Automatiser les extractions

Contenu

- Présentation de SQLAlchemy

- Connexion à PostgreSQL ou MySQL
- Lecture de données SQL
- Écriture de données
- Gestion des connexions
- Bonnes pratiques d'accès aux données

Mise en pratique

- Connexion à une base de données
- Exécution de requêtes SQL depuis Python
- Chargement des résultats dans Pandas

Module 6 – Manipuler les données avec Pandas (3h)

Objectifs

- Préparer les données pour l'analyse
- Réaliser des transformations de données

Contenu

- DataFrames Pandas
- Importation des données
- Nettoyage des données
- Filtrage
- Tri
- Agrégation
- Fusion de jeux de données
- Contrôle qualité des données

Mise en pratique

- Nettoyage d'un jeu de données réel
- Préparation d'un dataset analytique
- Production d'indicateurs

Module 7 – Analyse exploratoire et visualisation (2h)

Objectifs

- Réaliser une première analyse de données
- Présenter les résultats de manière visuelle

Contenu

- Analyse exploratoire des données
- Statistiques descriptives
- Introduction à Matplotlib
- Présentation de Seaborn
- Bonnes pratiques de visualisation

Mise en pratique

- Analyse d'un jeu de données
- Création de graphiques
- Interprétation des résultats

Module 8 – Atelier fil rouge : construire une chaîne d'analyse SQL et Python (2h)

Objectifs

- Mettre en œuvre les compétences acquises
- Réaliser un mini-projet de Data Analysis

Contenu

- Connexion à la base de données
- Extraction SQL

- Nettoyage avec Pandas
- Analyse des données
- Création de visualisations
- Présentation des résultats

Mise en pratique

- Réalisation complète d'un projet d'analyse de données
- Restitution des résultats
- Débriefing collectif

Notre référent handicap se tient à votre disposition au [01.71.19.70.30](tel:01.71.19.70.30) ou par mail à <mailto:referent.handicap@edugroupe.com> pour recueillir vos éventuels besoins d'aménagements, afin de vous offrir la meilleure expérience possible.