# Fiche du cours

**65 h**

**Titre :**

DS300 - Data Science (Python & SQL, de l'EDA au storytelling)

**Description :**

Parcours end-to-end de data science : cadrer une question métier, collecter/extraire (fichiers & SQL), nettoyer et qualifier les données, conduire une EDA (analyse exploratoire) avec Pandas/NumPy et visualisations (Matplotlib/Seaborn), appliquer les statistiques essentielles (inférence, tests d'hypothèses), construire un baseline modèle (scikit-learn – léger), mener une expérimentation A/B, et raconter les résultats (storytelling, dashboard/rapport). Focus sur qualité, éthique & confidentialité, traçabilité et reproductibilité.

**Objectifs :**

- Cadrer un problème métier et définir données, métriques, hypothèses.*
- Extraire/joindre des données avec SQL ; nettoyer, typer et documenter les jeux de données.*
- Réaliser une EDA rigoureuse et produire des graphiques lisibles et actionnables.*
- Appliquer les statistiques de base : intervalles de confiance, tests t/$\chi^2$, corrélations.*
- Construire un baseline prédictif (régression/logistique) et évaluer correctement (holdout/CV).*
- Concevoir et analyser une expérimentation (A/B) : puissance, effet, segmentation.*
- Produire un rapport/dash clair (insights, limites, recommandations) et garantir reproductibilité.

**Chapitres :**

1. Cadrage & pipeline DS : objectifs, métriques, risques de biais, éthique & RGPD (notions)*
2. Acquisition & SQL : schémas, jointures, agrégations, vues/CTE, extraction reproductible*
3. Qualité & nettoyage : valeurs manquantes, outliers, typage, dictionnaire de données*
4. EDA I : distributions, regroupements, corrélations, segmentations, dataviz efficace*
5. EDA II (time & catégorielles) : séries simples, saisonnalité, encodage, regroupements métiers*
6. Stats essentielles : IC, tests t/$\chi^2$, p-valeur vs taille d'effet, pièges courants*
7. Baseline ML (léger) : régression linéaire/logistique, métriques (RMSE, F1/AUC), validation*
8. Expérimentation : A/B testing (design, puissance), suivi, interprétation & limites*
9. Storytelling & livrables : rapport structuré, data viz finale, dashboard (notebook→rapport ou BI léger)*

10. Capstone : étude complète de bout en bout (SQL→EDA→stats→baseline→reco) + README & présentation

**À la fin :**

Vous saurez encadrer, explorer et analyser un jeu de données réel, produire des insights actionnables avec EDA + statistiques, bâtir un baseline crédible, tester une hypothèse (A/B) et livrer un rapport/storytelling convaincant avec recommandations — prêt à montrer en portfolio ou à passer à ML300/ML400. Vous mettrez en place une démarche reproductible (extraction SQL versionnée, notebook propre, suivi des sources et transformations). Vous saurez aussi expliciter les limites de l’analyse, proposer des pistes d’amélioration et préparer un petit tableau de bord pour le suivi des indicateurs clés.