

Abdel Nacer SAYAH

Data Engineer | Automatisation & Qualité des données

abdelnacer.sayah@gmail.com | +33695952444 | linkedin.com/in/abdelnacersayah | github.com/as-data

Compétences

Langages – Python, SQL, Bash

Base de données – PostgreSQL, Oracle

Environnements – Git, CI/CD, Linux

Frameworks / Data tools – Pandas, NumPy, Jupyter, Airflow, Spark, Dataiku

DataViz & BI – Power BI

Formations

Programme Grande Ecole de Commerce, ISC Paris

2017 – 2020

Master en Management des Systèmes d'Information

Modules : Programmation, Base de données, Data Mining, Machine Learning

Data Analyst Certification, Mines ParisTech PSL

2023

300 heures.

Modules : Programmation, Data Visualisation, Webscraping, Business Intelligence, Big Data

Data Upskilling

2025

200 heures.

Modules : Python avancé, SQL avancé, Git, CI/CD, Sreamlit, FASTapi, Spark, Airflow, Linux

Expériences Professionnelles

CGI

Data Engineer

Ministère de l'Economie DGFIP

Résumé : Responsable de la migration d'une base Access complexe vers PostgreSQL, automatisation de traitements comptables, et mise en place d'un cas d'usage IA dans Dataiku pour faciliter l'analyse métier.

- Migration d'une base multi-sources (10+ fichiers) vers PostgreSQL, réduisant les erreurs de jointure de 30%.
- Conception de pipelines semi-automatisés en Python/SQL, divisant par deux le temps de traitement des données.
- Mise en place de workflows ETL sous Dataiku pour la traçabilité et la validation métier.
- Audit de cohérence des rapports DSI : suppression de sources obsolètes, fiabilisation des extractions critiques.
- Déploiement d'un démonstrateur IA dans Dataiku (NLP + classification) réduisant le temps d'analyse des données.

Data Analyst

Ministère de la Justice

Résumé : Chargé de la qualité des données dans le cadre d'une migration applicative à l'échelle nationale. Mise en place d'outils d'audit et de suivi métier pour fiabiliser le socle de données.

- Audit de 400 000+ lignes : détection de doublons, nettoyage de formats, recommandations de traitement d'anomalies.
- Création d'un pipeline Python pour tracer les rejets de migration, réduisant le taux d'anomalies non expliquées de 40%.
- Conception d'un tableau de bord Power BI utilisé hebdomadairement par les référents métier pour le pilotage du redressement de données.
- Rédaction d'un dictionnaire de données validé en comité projet.

Consultant AMOA

Office National des Forêts

Résumé : Intermédiaire entre la DSI et les métiers pour cadrer, spécifier et suivre les évolutions fonctionnelles d'un SI en TMA. Forte exposition aux cycles de validation et à l'accompagnement utilisateur.

- Rédaction et mise à jour de spécifications fonctionnelles détaillées.
- Exécution et validation de 80+ cas de test en recette (SQL Oracle).
- Suivi de tickets JIRA : taux de clôture >90%
- Animation d'ateliers utilisateurs pour la conduite du changement.

Projets

chat_clip_classifier, Classifier automatiquement les vidéos d'un chat

2024

Résumé : Traitement de 358 vidéos (15h) en 1 157 clips classés via un pipeline Python

- Objectif : automatiser le dérushage de contenus TikTok pour réduire le temps de montage de 80%. Classification zero-shot basée sur 12 descriptions comportementales.

Stack : Python, OpenCV, CLIP (modèle IA), PyTorch, ffmpeg, Git, Bash