

ANAS JEBALI

DATA ANALYST / DATA ENGINEER

BI, pipelines de données et IA appliquée

jebalianas00@gmail.com • +33 6 36 21 59 16 • Montpellier, France

[LinkedIn](#) • [GitHub](#) • [Portfolio](#)

PROFIL

Ingénieur de la donnée de formation, je me suis construit sur le terrain — en transformant des données brutes en outils que les directions utilisent vraiment. Ce qui m'anime : faire le lien entre la technique et la décision, et livrer des solutions qui tiennent en production. **Disponible en CDI à partir de septembre 2026.**

EXPÉRIENCES PROFESSIONNELLES

Data Analyst / Data Engineer — Alternance | Groupe Tressol-Chabrier • Montpellier *Sept. 2024 – Sept. 2026*

- **Pipeline NLP de classification multi-label des avis clients Google My Business** : CamemBERT fine-tuné sur une taxonomie métier à 2 axes (10 thèmes, type de véhicule). F1 de 0,83 à 0,93 sur les thèmes principaux — industrialisé sur 38 000 avis.
- **Application Streamlit NL2SQL** : exploration du référentiel client via langage naturel (API Mistral Large, Python, pyodbc, Plotly) — interrogation sans SQL pour les équipes métier.
- **Data Warehouse Silver Layer sur SQL Server** : archivage optimisé de 19 tables Autolink DMS (6,6M+ lignes), modélisation en étoile, nettoyage et fusion de données multi-sources.
- **Unification des données clients de 3 systèmes DMS hétérogènes** : data warehouse SQL Server, modélisation et réconciliation des référentiels.
- **Audit qualité sur un data warehouse multi-source et multicouche (Staging → Silver → Gold)** : diagnostic et correction d'un défaut de consolidation sur un référentiel véhicules, avec récupération de 4 262 valeurs perdues par la logique d'arbitrage inter-sources.
- **Pipeline d'extraction Google My Business (API Partoo)** : conteneurisé Docker, versionné Azure DevOps, contrôles qualité hebdomadaires.
- **Automatisation SQL Server → CRM OpenFlex** : temps de traitement ramené de 45 à 4 minutes par exécution (−91 %).
- **Dashboard Power BI de pilotage commercial et après-vente** : DAX, Power Query, hiérarchies plaque/marque/ville — utilisé par le CTO et les directions Marketing, VN et VO, avec revue périodique des indicateurs par métier.
- **Formation de l'équipe Tech & Data** : extraction multi-sources et optimisation SQL.

Data Analyst / BI — Freelance | JobAppeal • Montpellier *Oct. 2025 – Mars 2026*

- **Pipeline ELT en production** : extraction des échanges clients (Gmail API), chargement BigQuery, orchestration Cloud Scheduler, restitution Power BI Service — en remplacement d'un suivi manuel. Mission cadrée en direct avec la direction.

Data Analyst — Stage | CHU de Montpellier *Mai – Juil. 2024*

- **Évaluation d'un modèle génératif en contexte médical** : ROUGE-4 et BERTScore (Python, Hugging Face) — conclusions restituées aux médecins et à l'encadrement.

PROJETS

- **CycleGAN bidirectionnel — Computer Vision (Caltech Camera Traps)** : domain adaptation jour/nuit, réduction de 74 % du gap d'accuracy. Stack : PyTorch, OpenCV.
- **GeoLifeCLEF 2025 — Prédiction multi-label d'espèces végétales** : pipeline TabularMLP sur serveur GPU partagé, explicabilité (Grad-CAM, SHAP), ablation de modalités. Stack : PyTorch, Scikit-learn.

FORMATION

Master MIASHS — Maths & Informatique Appliquées | Univ. Paul-Valéry, Montpellier 3 *2024 – 2026*

Modules clés : Deep Learning, NLP, Séries temporelles (ARIMA, GARCH), Big Data (Spark), Graphes, Modèles mixtes.

Licence MIASHS — Maths & Informatique Appliquées | Univ. Paul-Valéry, Montpellier 3 *2020 – 2024*

COMPÉTENCES TECHNIQUES

Langages & bases Python (avancé) • SQL (avancé) • R • Bash • MongoDB

Data Engineering ETL/ELT • Pipelines de données • API REST • Data Warehouse • Schéma en étoile • Architecture Medallion • PySpark

BI & Dataviz Power BI (avancé) • DAX • Power Query • Power BI Service • Streamlit • Tableau

ML / NLP Scikit-learn • PyTorch • Hugging Face • CamemBERT • NLP • LLM • RAG • Computer Vision

Cloud & DevOps GCP (BigQuery, Cloud Scheduler, Compute Engine, Cloud Functions) • Azure DevOps • Docker • Git

Statistiques & méthodes Tests d'hypothèses • Séries temporelles • Modèles mixtes • Inférence bayésienne • Agile • JIRA

CERTIFICATIONS & LANGUES

Generative AI with Diffusion Models — NVIDIA Deep Learning Institute (2025) | **Langues** : Français (natif) • Anglais (C1, professionnel) • Arabe (courant)