Imane EL Missaoui

DATA ENGINEER | PYTHON, SQL, AIRFLOW & DATA QUALITY

Paris Area, France | (+33) 6 98 75 95 99 | imaneelmissaoui23@gmail.com | linkedin.com/in/imane-el-missaoui | Portfolio

Data Engineer with 2+ years of combined experience designing batch pipelines, integrating heterogeneous sources, modeling data and improving data quality. Professional experience with Python, SQL, Airflow, PostgreSQL, MongoDB, Git and Linux; dbt and Docker applied in an end-to-end portfolio project. Skilled at turning complex and sensitive data into reliable, documented, analysis-ready datasets.

PROFESSIONAL EXPERIENCE

Medical Imaging Data Scientist (Data Engineering focus) | LYSARC, Créteil, France

04/2025 - 04/2026 | *Multicenter PET/CT batch pipeline*

- Designed a Python batch pipeline integrating heterogeneous DICOM and CSV sources, with NIfTI standardization, mask preparation and structured outputs.
- Automated feature extraction with PyRadiomics and consolidated results into analysis-ready tables for downstream teams.
- Implemented quality checks on metadata, dimensions, spacing, orientation and image-mask consistency to detect non-compliant data.
- Documented business rules, pipeline architecture and run procedures to support reproducibility, traceability and handover.

Data Science Apprentice & Medical Imaging Engineer | Sorbonne University Computer Science Laboratory, Paris, France

12/2023 - 09/2024 | *Databases and workflow automation*

- Designed and administered MongoDB dev/test/prod environments to centralize metadata for an endoscopic capsule project with 10,000+ images.
- Defined schemas, formats and quality rules; automated processing and versioned changes with Git on Linux.
- Built a Python and Airflow workflow processing 500 Hz PPG, ECG and temperature signals and serving KPIs through Streamlit.
- Produced technical documentation and worked with researchers, clinicians and engineers to make data reusable.

Data Science Intern & Medical Imaging Engineer | Lille Neuroscience & Cognition, Lille, France

03/2023 - 08/2023 | *MRI data pipelines*

- Prepared and quality-checked 2,000+ multicohort 3D T1 MRI scans for a Parkinson's disease patient stratification project.
- Built reproducible loading, preprocessing and feature-extraction workflows using Python, DIPY, Jupyter and Shell scripts.
- Standardized intermediate datasets and documented processing steps to improve analysis reuse and reproducibility.

DATA ENGINEERING PROJECT - PORTFOLIO

Clinical trials data pipeline | *Python, PostgreSQL, dbt, Docker Compose, Metabase*

- Ingested JSON data from the ClinicalTrials.gov API into PostgreSQL and organized it into bronze, silver and gold layers.
- Built SQL transformations with dbt and a star schema to deliver reliable, analysis-ready tables.
- Orchestrated the pipeline in Python, containerized it with Docker Compose, exposed results in Metabase and documented the full workflow.

TECHNICAL SKILLS

Languages: SQL, Python | R
Pipelines & orchestration: ETL/ELT, REST APIs/JSON, Airflow, dbt, batch pipelines
Databases & modeling: PostgreSQL, MongoDB, Neo4j | relational, star schema, medallion layers
Development & deployment: Git, Linux, Docker Compose, Jupyter
Quality & operations: validation checks, traceability, reproducibility, monitoring, documentation
Cloud & distributed: AWS, Spark, Databricks (foundations being strengthened)
Reporting & formats: Metabase, Streamlit | DICOM, NIfTI, PyRadiomics

EDUCATION & LANGUAGES

2022 - 2024: Master's Degree in Health Data Science - University of Lille, France
2020 - 2022: Bachelor's Degree in Physics and Chemistry - Université Grenoble Alpes, France
2019 - 2020: First year of the Engineering Cycle in Biomaterials Engineering - Grenoble INP | Admitted through a competitive entrance exam after CPGE Mathematics-Physics
2017 - 2019: CPGE MPSI/MP intensive Mathematics-Physics program, Excellence Scholarship - Lycée d'Excellence de Benguerir
2025 - 2026: Data Engineering on AWS learning path - LinkedIn Learning
Languages: French: fluent | English: C2