# Samuel Okotako

Data Engineer

Catania, Italie | +39 351 092 0026 | francissamuel840@gmail.com
Portfolio · GitHub · LinkedIn

## PROFIL PROFESSIONNEL

Titulaire du diplôme ITS de technicien supérieur en Data Management, je développe des pipelines Python et SQL pour l'ingestion, les transformations incrémentales et la qualité des données. Mes projets utilisent DuckDB et dbt pour les modèles analytiques, ainsi que PySpark sur Databricks pour l'acquisition de données brutes. J'applique des tests et des contrôles de traçabilité.

## EXPÉRIENCE PROFESSIONNELLE

**Spindox | Data Management Intern** Catania, Italie | juin–octobre 2026

- Approfondi les chargements incrémentaux, SQL relationnel et Databricks dans le cadre de la formation technique du stage.
- Développé le projet individuel *Data Privacy Governance* sur des données synthétiques : inventaire de 109 colonnes PostgreSQL et transformations de masquage et de pseudonymisation.
- Vérifié l'intégrité et les relations entre 20 tables protégées ; la suite de tests documentée compte 182 tests réussis.

## PROJETS SÉLECTIONNÉS

**Pipeline e-commerce incrémental** Projet personnel | Python, DuckDB, dbt | Étude de cas

- Construit sur des commandes synthétiques un pipeline Python, DuckDB et dbt ; les upserts déterministes gèrent 10 insertions tardives et 5 corrections sans doublons.
- Validé 5 010 lignes courantes, des rejouements idempotents et l'équivalence avec un rechargement complet ; les indicateurs sont présentés dans Streamlit.

**Inside Airbnb sur Databricks** Projet ITS en groupe | PySpark, Delta

- Coordonné le groupe et écrit l'ingestion Bronze de quatre CSV d'Amsterdam et Rio, avec les métadonnées de ville et d'horodatage.
- Stocké les données brutes dans des tables Delta ; le projet collectif comprend des transformations Silver et des vues SQL Gold pour l'analyse comparative.

## COMPÉTENCES TECHNIQUES

**Ingénierie des données** Python, SQL, ETL/ELT, dbt, DuckDB, Databricks

**Plateformes de données** PostgreSQL, PySpark, Delta Lake, architecture Bronze/Silver/Gold

**Cloud & DevOps** AWS (formation ITS), Docker, Git

**Qualité des données** tests automatisés, validation, intégrité, pipelines incrémentaux

## LANGUES

Anglais – Langue maternelle | Italien – C1 | Français – B1

## FORMATION

**ITS Steve Jobs Academy** 2024–2026
Diplôme de technicien supérieur en Data Management (EQF 5) – 110/110

**Università degli Studi di Catania**
Licence en informatique (L-31) – en cours

**I.T. Archimede** 2018–2023
Diplôme de lycée technique en électronique et électrotechnique – spécialisation automatisation

J'autorise le traitement des données personnelles contenues dans ce CV à des fins de recrutement, conformément au règlement (UE) 2016/679.