

Geovany Batista Polo Laguerre

Data Engineer · Data Analyst

lgeobatpo98@gmail.com · linkedin.com/in/geobatpo07 · github.com/Geobatpo07 · Île-de-France

PROFILE

Data engineer and data analyst, Microsoft Fabric certified (DP-700, DP-600). From 2023 to 2026, ETL pipelines, lakehouse architectures and Power BI dashboards on healthcare data, with steady attention to data quality. Master's degree in applied mathematics.

EXPERIENCE

Data Analyst · Solutions S.A. — Jul 2023 – Aug 2026 · Remote · Healthcare sector

- Analytics architecture on Microsoft Fabric (Lakehouse, Data Warehouse).
- Power BI dashboards and KPI tracking; helped raise the share of client sites above 95% performance from about 20% to about 60%.
- ETL pipelines and healthcare data quality, including machine-learning-based deduplication.
- Mapping with QGIS and Power BI.
- SQL Server, MySQL and PostgreSQL databases.

Data Engineer (volunteer) · AYITISTATS — Aug 2026 – present

- Unifying statistical data for a Google Data Commons instance on Google Cloud (GCP).

Lecturer, Mathematics for Computer Science · Université Quisqueya — Sep 2025 – present

- Mathematics for computer science, taught with a focus on its uses in data science, software engineering and web technologies.

Teaching Assistant, Data Science & AI · Université Quisqueya — May 2025 – present

- Supporting a hands-on curriculum run with Akademi and aligned with the Flatiron School immersive program.
- Data wrangling, exploratory analysis, model building and deployment with Python, Pandas, NumPy, Scikit-learn and Jupyter.
- Code reviews, Q&A sessions and contributions to curriculum refinement.

PROJECTS

DataHut-DuckHouse · github.com/Geobatpo07/datahut-duckhouse

A multi-tenant analytics platform, local or cloud: Arrow Flight ingestion, hybrid DuckDB and Iceberg storage (MinIO/S3), dbt transformations and Trino queries. Stack: Python, DuckDB, Apache Iceberg, Arrow Flight, dbt, Trino, MinIO.

Nyansa · github.com/Geobatpo07/nyansa

A self-hosted AI second brain on a private server: seven Docker Compose services (local LLMs with Ollama, Qdrant vector store, n8n workflows) wired to an Obsidian vault. Built on n8n's self-hosted AI starter kit. Stack: Docker Compose, n8n, Ollama, Qdrant, PostgreSQL, Open WebUI.

Scientific Assistant · github.com/Geobatpo07/scientific-assistant

A local multi-agent system: eight specialized agents (planning, symbolic mathematics, numerical methods, statistics, literature, review, writing, memory) and hybrid vector and keyword RAG. Produces executable Python code and citations. Stack: Python, LangChain, Ollama, ChromaDB, RAG.

SKILLS

Languages: Python, SQL, R

Databases: SQL Server, PostgreSQL, MySQL, DuckDB

Data engineering: Microsoft Fabric (Lakehouse, Data Warehouse), ETL, dbt, Apache Iceberg, Trino, Arrow Flight, Spark SQL, Google Cloud (GCP)

BI & visualization: Power BI, QGIS, Shiny

Machine learning & AI: Scikit-learn, Pandas, NumPy, RAG, Multi-agent systems, Local LLMs (Ollama)

DevOps & tooling: Docker, Linux, Git

Mathematics & scientific computing: Statistics & probability, Compartmental modelling

EDUCATION

Master's in Mathematics and Applications, MOAD track · Université des Antilles · 2024 – 2026

MicroMasters in Statistics and Data Science · MIT (MITx) · 2024

Bachelor's degree in Mathematics · École Normale Supérieure, Université d'État d'Haïti · 2021 – 2022

Civil Engineering degree · Faculté des Sciences, Université d'État d'Haïti · 2017 – 2022

Certifications: Microsoft Certified: Fabric Data Engineer Associate (DP-700) · Microsoft Certified: Fabric Analytics Engineer Associate (DP-600) · Apache Spark SQL for Data Analysts · Google Data Analytics Professional Certificate · Google Business Intelligence Professional Certificate