

Geovany Batista Polo Laguerre

Data Engineer · Data Analyst

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

PROFIL

Data engineer et data analyst certifié Microsoft Fabric (DP-700, DP-600). De 2023 à 2026, pipelines ETL, architectures Lakehouse et tableaux de bord Power BI sur des données de santé, avec une attention constante à leur qualité. Master de mathématiques appliquées.

EXPÉRIENCE

Data Analyst · Solutions S.A. — juil. 2023 – août 2026 · À distance · Secteur santé

- Architecture analytique sur Microsoft Fabric (Lakehouse, Data Warehouse).
- Tableaux de bord Power BI et suivi des indicateurs (KPI) ; contribution à la hausse de la part des sites clients au-dessus de 95 % de performance, d'environ 20 % à environ 60 %.
- Pipelines ETL et qualité des données de santé, dont une déduplication par machine learning.
- Cartographie avec QGIS et Power BI.
- Bases de données SQL Server, MySQL et PostgreSQL.

Data Engineer (bénévole) · AYITISTATS — août 2026 – aujourd'hui

- Unification de données statistiques pour une instance Google Data Commons sur Google Cloud (GCP).

Chargé de cours, mathématiques pour informaticiens · Université Quisqueya — sept. 2025 – aujourd'hui

- Cours de mathématiques pour informaticiens, orientés vers leurs usages en data science, en génie logiciel et en technologies web.

Assistant d'enseignement, Data Science & IA · Université Quisqueya — mai 2025 – aujourd'hui

- Accompagnement d'un cursus pratique mené avec Akademi et aligné sur le programme intensif de la Flatiron School.
- Préparation des données, analyse exploratoire, construction et déploiement de modèles avec Python, Pandas, NumPy, Scikit-learn et Jupyter.
- Revues de code, séances de questions-réponses et contribution à l'amélioration du programme.

PROJETS

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

Plateforme analytique multi-tenant, locale ou cloud : ingestion par Arrow Flight, stockage hybride DuckDB et Iceberg (MinIO/S3), transformations dbt et requêtes Trino. Technologies : Python, DuckDB, Apache Iceberg, Arrow Flight, dbt, Trino, MinIO.

Nyansa · github.com/Geobatpo07/nyansa

Second cerveau IA auto-hébergé sur un serveur privé : sept services Docker Compose (LLM locaux avec Ollama, base vectorielle Qdrant, workflows n8n) reliés à un coffre Obsidian. Construit à partir du self-hosted AI starter kit de n8n. Technologies : Docker Compose, n8n, Ollama, Qdrant, PostgreSQL, Open WebUI.

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

Système multi-agents exécuté en local : huit agents spécialisés (planification, mathématiques symboliques, méthodes numériques, statistiques, bibliographie, relecture, rédaction, mémoire) et un RAG hybride vectoriel et lexical. Produit du code Python exécutable et des citations. Technologies : Python, LangChain, Ollama, ChromaDB, RAG.

COMPÉTENCES

Langages : Python, SQL, R

Bases de données : 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 et visualisation : Power BI, QGIS, Shiny

Machine learning et IA : Scikit-learn, Pandas, NumPy, RAG, Systèmes multi-agents, LLM locaux (Ollama)

DevOps et outillage : Docker, Linux, Git

Mathématiques et calcul scientifique : Statistiques et probabilités, Modélisation compartimentale

FORMATION

Master Mathématiques et applications, parcours MOAD · Université des Antilles · 2024 – 2026

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

Licence de mathématiques · École Normale Supérieure, Université d'État d'Haïti · 2021 – 2022

Diplôme d'ingénieur civil · 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