Guilherme Yuji Marcolongo Adania

Data Engineering | Data Analytics | BI | Automation
São Paulo, Brazil | +55 (11) 99334-5225 | gyuji1904@gmail.com | linkedin.com/in/guilherme-yuji | github.com/GuiYuuji

PROFESSIONAL SUMMARY

Data and BI professional with hands-on experience across data engineering, data quality, Python automation, SQL, REST APIs, Big Data, and data governance in enterprise environments. Experienced in designing and shipping internal data products end to end — from API integrations and validation engines to automated monitoring systems and reporting solutions — as well as building full-stack systems for external clients as an independent developer. Strong analytical profile with a practical Data Engineering mindset, focused on scalable, accurate, and maintainable solutions.

PROFESSIONAL EXPERIENCE

Engineering Brasil — Junior Data Analyst | Vivo Project | Mar 2026 – Present

- Lead the Engineering Brasil initiative for creating AI agents and automation solutions for Vivo, working with Python, LangGraph, and multi-agent architectures to automate business workflows (initiative in early stage).
- Work in the Data Curation workstream for the Vivo project, supporting data organization, metadata, cataloging, governance, and quality processes.
- Designed and built a metadata briefing validation system integrated with Informatica EDC via REST APIs, using parallel HTTP requests and thread pooling. Validates assets, databases, tables, fields, descriptions, business names, and additional metadata attributes in a single run.
- **Impact:** briefings above 2,000 rows were previously unfeasible or took hours on the legacy process; briefings above 15,000 rows now complete in minutes — approximately 30x faster for single validations and 85x faster for continuous validation cycles.
- Built an automated monitoring system covering ~70 multi-table projects that detects volumetry deviations, data load delays, and SLA breaches, replacing a fully manual daily check performed by two analysts.
- **Impact:** ~60 incidents detected in the first 4 months of operation; daily validation time reduced from 3–4 hours to under 1 hour (~70%).
- Extended the monitoring system with automated visual dashboard validation to catch silent refresh failures — dashboards that update without error but render blank or corrupted. Runs on a fixed intra-day schedule, exports the current dashboard state to PDF, converts it to PNG, and compares it against a reference image using OpenCV to flag anomalies. Built after a client-reported broken dashboard; has since flagged 2 real failures.
- Perform Data Quality analysis, validation, monitoring, and troubleshooting across enterprise data processes, investigating inconsistencies in records, metadata, and technical documentation.
- Work with data cataloging, traceability, taxonomy, and governance processes to improve the organization and reliability of data assets.
- Collaborate with technical and business stakeholders to standardize processes, document requirements, and improve data quality across teams.
- Work with medallion architecture (bronze / silver / gold) and data domain organization within the project's data platform.
- Operate data workloads on GCP virtual machines within the enterprise environment.
- **Enterprise stack:** Python, SQL, Databricks, PySpark, Hive, Teradata, DBeaver, Power BI, Informatica EDC, REST APIs, GCP, Azure DevOps.

Cellere Soluções em TIC — Data Analyst | Vivo Project | May 2025 – Dec 2025

- Developed Python automations that reduced recurring manual work for analysts and improved operational efficiency.
- Created and maintained internal systems and data integrations between platforms through REST APIs and automated workflows.
- Performed data analysis and curation in Big Data environments, validating tables, fields, structures, and metadata at scale.
- Wrote SQL queries in distributed Hive/Tez environments over partitioned tables for data validation, reconciliation, and cross-source analysis.
- Reviewed technical data briefings and translated requirements between technical and business teams, supporting data governance and documentation.

Independent Developer — Freelance Software Development | 2022 – Present

- Delivered custom systems end to end for approximately 15 small and mid-sized business clients, including auto repair shops, optical stores, and similar service companies.

- Handled the full cycle: requirements gathering with non-technical clients, system design, development, deployment, and post-delivery support.
- Built web systems and internal management tools using PHP, Next.js, React, and TypeScript, containerized with Docker and deployed to VPS environments.

CORE TECHNICAL SKILLS

Programming & Automation: Python, SQL, JavaScript, TypeScript, PHP, REST APIs, scripting, process automation, parallel processing / thread pooling, Selenium, OpenCV, openpyxl

Data Engineering: ETL/ELT, data pipelines, data transformation, data integration, data quality, QA/QC, medallion architecture, data modeling, partitioning, troubleshooting, root-cause analysis

Data Platforms & Processing: Databricks, PySpark, pandas, Hive/Tez, Teradata, Big Data, GCP, Informatica EDC, PostgreSQL, Supabase

Data Formats: Parquet, JSON, CSV, XML

BI & Reporting: Power BI, dashboard development, data visualization, recurring reporting workflows

AI & Agents: LangGraph, multi-agent architectures, OpenAI APIs, LLM-based automation

Web & Frameworks: Next.js, React, Flask, FastAPI, Streamlit, HTML/CSS

DevOps & Tools: Docker, Git, Azure DevOps, VPS deployment, DBeaver, Figma

Data Governance: Data Quality, metadata management, cataloging, taxonomy, traceability, governance, technical documentation

EDUCATION

Bachelor's Degree in Systems Analysis and Development — Universidade Presbiteriana Mackenzie | Feb 2026 – Dec 2027
Technical Degree in Systems Development — ETEC | Feb 2020 – Dec 2022
Previously: Systems Analysis and Development — FATEC (Feb 2023 – Dec 2025, transferred to Mackenzie)

CERTIFICATIONS & TRAINING

CS50x: Introduction to Computer Science — Harvard University | Mar 2026
Data Engineering Foundations Professional Certificate — Astronomer | Mar 2026
OpenAI & Python: APIs and Chatbots — Alura | Nov 2025
Data Engineering with Databricks — Alura | Jul 2025
Data Science — Alura | Jun 2025
Machine Learning in Practice — Alura | May 2025

LANGUAGES

Portuguese — Native | English — C2 (EF SET)