Michael Ozeh Okang

Data Engineer | Data Pipelines · Data Quality & Governance · Analytics Engineering

Vienna, Austria · +43 664 994 75957 · okang.michael@gmail.com · https://okangmichael.com · LinkedIn · GitHub

PROFESSIONAL SUMMARY

Data Engineer with over three years of commercial business analytics experience across a 30-vehicle FMCG distribution network, analysing operational costs, revenue, payroll and fleet performance to inform business decisions. Independently develops governed batch and streaming pipelines, analytical models and reporting-ready datasets, with validation, lineage and recoverability controls. Built a lakehouse with eight analytical marts and 38 passing automated data-quality tests. Based in Vienna; English (native), German (B1).

TECHNICAL SKILLS

- **Languages & analytics:** Python, SQL, PySpark, Spark SQL, Power BI, DAX
- **Excel & reporting:** Power Query (clean/merge), PivotTables (KPIs), XLOOKUP/VLOOKUP (reconcile), dashboards
- **Data engineering:** Databricks, Delta Lake, dbt, medallion, ETL/ELT, incremental loads, star schema
- **Streaming and integration:** Kafka, Confluent Cloud, Snowpipe, REST APIs, schema validation
- **Platforms and databases:** Snowflake, BigQuery, SQL Server, MySQL, PostgreSQL
- **Orchestration and governance:** Airflow, Unity Catalog, lineage, provenance, automated testing, Git, Docker
- **Foundational knowledge:** CDC, SCD Types 1/2, RBAC, observability, Databricks Workflows

PROFESSIONAL EXPERIENCE

Business Data Analyst & Logistics Operations Partner | Self-employed, Lagos, Nigeria **Apr 2021 – Nov 2024**

- Maintained and analysed structured Excel data covering trip records, routes, driver rosters, payroll, vehicle maintenance, operating costs and revenue across a 30-vehicle FMCG distribution network.
- Performed KPI reporting and route-profitability analysis to assess vehicle utilisation, driver performance and cost drivers, informing budgeting, resource allocation and operational efficiency.
- Processed driver payroll from trip and roster data, reconciled transportation charges, and issued B2B invoices to support accurate financial reporting and cost control.
- Used analytical insights to plan fleet operations and coordinate drivers, manufacturers, suppliers, wholesalers and customers as an independent B2B service partner to TradeDepot and AAVA Brands Limited.

Service Station Associate | Shell, Hof bei Salzburg, Austria **Feb 2025 – Feb 2026**

- Handled point-of-sale transactions and cash reconciliation accurately under time pressure; monitored inventory and reported discrepancies.

INDEPENDENT DATA ENGINEERING PROJECTS

Independent Data Engineer and Analytics Developer | Vienna, Austria **Jun 2025 – Present**

Project Alpine: Real-Time Market Data Lakehouse | *Kafka, PySpark, Databricks, Delta Lake, dbt*

- Built a streaming lakehouse ingesting Kraken and Coinbase market events through Python producers and dedicated Confluent Cloud topics into Bronze, Silver and Gold Delta tables.
- Delivered 3 governed Silver branches, a canonical market-price fact and 8 dbt Gold marts, validated by 38 passing automated data-quality tests.
- Protected data quality with source-contract validation, provenance tracking, quarantine handling and checkpointed writes; managed schemas and access through Unity Catalog.

Vienna Transit Intelligence Platform | *Python, BigQuery, dbt, Airflow*

- Built a scheduled batch ELT pipeline combining Wiener Linien data with Open-Meteo weather for 150+ U-Bahn platform stops (U1–U6) and six weather stations, matched by Haversine distance.
- Kept history traceable and recoverable with append-only BigQuery tables, incremental dbt models, data-quality tests and Airflow backfills.
- Replaced the weather source (GeoSphere to Open-Meteo) when it no longer met project scope, without changing downstream processing.

Cart Abandonment Engine | *Kafka, Snowflake, Snowpipe, dbt, Airflow, Docker*

- Streamed click events through Confluent Cloud into Snowflake and used dbt window functions (gaps-and-islands) to flag one abandoned cart per session after 15 minutes of inactivity.
- Automated hourly dbt runs and validation in Dockerized Airflow, publishing a documented, reverse-ETL-ready mart.

Analytics projects | *SQL Server, Python, Power BI, DAX, star schema*

- Chocolate Sales Intelligence: audited ~200,000 sales records and recovered ~1,900 legitimate orders excluded by a faulty tolerance rule; modelled a Power BI star schema.
- Cultural Heritage and Sustainability Analytics: validated 15,000 records across 50 sites, corrected an invalid fixed-site assumption and built a dimensional model at the correct grain.

Planned (not implemented): Project Atlas, a version-aware GraphRAG concept for Austrian legal retrieval. Architecture designed with governed retrieval in mind; development has not started.

EDUCATION

MSc Sustainability Assessment and Resource Management (part-time) **Sep 2026 – Present**
Hochschule Campus Wien

MSc Applied Geoinformatics **Mar 2025 – Present**
University of Salzburg. 34 ECTS completed

BSc Urban and Regional Planning **Dec 2006 – Jan 2013**
Olabisi Onabanjo University. Second Class Upper

CERTIFICATIONS

Mastering Advanced SQL Queries; SQL Foundations; Data Manipulation and Transactions in SQL Server; Power BI Essentials and Advanced Techniques (Coursera, Microsoft and Packt, all Jul 2026)