FRANK SARFO

Data Engineer & Scientist

Goldammerstr. 47, 12351, Berlin

sarfof06@gmail.com

+49 152 1687 1042

https://www.linkedin.com/in/frank-sarfo-13ab96253

https://github.com/bulla-01

Core Competencies

- Python (pandas, requests, asyncio, FastAPI)
- SQL (PostgreSQL, MySQL, BigQuery)
- Airflow (ETL orchestration)
- AWS (S3, Lambda, EC2, RDS)
- Data Lakes & Warehousing
- Large Language Models (OpenAI APIs, prompt design)
- Risk & Fraud Data Modeling
- CI/CD & Git Workflows
- Low-code/No-code tools (Retool, Streamlit)
- Communication & Project Leadership

EDUCATION

Master of Science: Data Science
Arden University - Berlin, Berlin Campus, (Distinction)

Bachelor of Science: BSc. Computing with Accounting
University for Development Studies - Tamale, Ghana
September 2011 – July 2016

Certification

AWS Cloud Certified – Professional Data Engineer *(In Progress)*

Languages

- English – Fluent (written and spoken)
- German – A2, in progress

PROFESSIONAL SUMMARY

Data Engineer & Scientist with over 4 years of experience designing scalable data pipelines, automating workflows with Python, and building cloud-based architectures in AWS. Passionate about using data to power sustainability and climate solutions. Proven ability to collaborate cross-functionally and lead data infrastructure initiatives from concept to deployment. Currently seeking to contribute technical skills and climate-driven motivation to CEEZER's mission of accelerating enterprise decarbonization through the voluntary carbon market.

EXPERIENCE

Data Scientist and Data Engineer | Access Bank Ghana Plc | March 2018 – April 2024

- Developed a hybrid ML system (LSTM + Random Forest) to detect mobile money fraud using sequential and tabular data.
- Built a real-time FastAPI backend with a frontend dashboard (Chart.js + Tailwind) for monitoring fraud risk, user behavior, and financial activity.
- Managed ingestion of large transaction datasets and model deployment with latency constraints for prediction APIs.
- Designed and maintained data pipelines that handled millions of transactional records using SQL and Python-based ETL tools.
- Applied OpenAI LLMs for data validation, enrichment, and natural language querying of fraud logs.
- Stored and queried data from an S3-backed data lake with SQL-based transformation logic.
- Engineered secure backend systems and automated regulatory reports using Python and SQL.
- Collaborated with cross-functional teams to integrate third-party APIs and internal tools.

MSc Research Projects, Arden University | Berlin | 2024 – 2025

- Designed a cloud-first infrastructure using AWS for ingesting and analyzing real-time medical device data.
- Built scalable pipelines with Python and orchestrated with Airflow to eliminate manual spreadsheet tracking.
- Proposed architecture to handle structured and unstructured data in compliance with healthcare regulations.
- Built interactive dashboards enabling clinicians to access data insights via web-based interfaces.