

Samuel Shaibu

Vilnius, Lithuania | +37062137540 | shaibu.samuel@yahoo.com
linkedin.com/in/samuel-shaibu | github.com/samietex | samietex.github.io

Data Science & Data Engineering

Profile

Data scientist and engineer with 4+ years of experience building robust data pipelines, applying statistical and ML models, and delivering production-grade data solutions. Strong in Python and SQL, with hands-on experience designing and maintaining ETL workflows, integrating data from diverse sources, and developing data science tools. Experienced applying statistical modelling techniques, including pricing, perception, and predictive models, to support decision-making. Combines solid software engineering practices (Git, CI/CD, testing) with a proactive, curious approach to understanding functional data domains and collaborating with stakeholders.

Work Experience

Syno International

Data Scientist

Vilnius, Lithuania

Jul 2025 – Present

- Designed and developed ETL pipelines for data ingestion, cleansing, and transformation across bronze and silver layers, ensuring data quality and integrity at each stage.
- Applied statistical and ML modelling (Gabor-Granger pricing models, CVI/CBI computations, and brand perception modelling) supporting 10+ international analytical studies with improved predictive accuracy.
- Productionised a legacy data integration pipeline: modernised a large Python codebase, added multi-format ingestion (CSV/Excel, SFTP/FTPS), AWS S3 delivery, and automated scheduling with logging and monitoring.
- Implemented privacy-safe data handling including cryptographic hashing, file encryption, MongoDB-based quarantine, and country-specific compliance rules, reducing operational risk.
- Developed and deployed AI-powered tools with Python, Streamlit, and LangChain, reducing manual workload by 40%.
- Implemented pipeline orchestration using EventBridge Scheduler and Step Functions with parallel upstream execution and dependency-managed downstream refresh; infrastructure provisioned with Terraform.

Syno International

Data Analyst

Vilnius, Lithuania

Sep 2024 – Jun 2025

- Built and maintained brand tracker dashboards in Power BI and SQL for global clients across multiple markets.
- Performed MaxDiff and Van Westendorp pricing analyses, improving clients' pricing strategy precision by 15–20%.
- Conducted key driver analysis and mood score modelling to identify factors influencing brand preference.
- Wrote SQL logic for CVI/CBI modelling, ensuring 95%+ data accuracy in insights delivered to decision-makers.
- Collaborated with stakeholders to gather requirements and deliver analytical outputs with clear documentation.

DataCamp

Technical Writer (Data Science)

Remote

Apr 2024 – Jul 2025

- Wrote and published 18+ peer-reviewed articles and tutorials on Python, SQL, statistical modelling, and data science.

GOMYCODE

Data Science Instructor

Lagos, Nigeria

Aug 2023 – Feb 2024

- Taught Python, SQL, machine learning, and statistical modelling to 40+ students; led end-to-end data pipeline projects.

Freelance

Business Intelligence Specialist / Data Scientist

Remote

Nov 2020 – Aug 2024

- Designed data solutions for clients across energy, finance, and edtech; built ETL processes with SSIS and SQL and Power BI dashboards for operational KPI reporting.

Selected Project

Production ETL Pipeline: Survey & Ad-Awareness Analytics

2026

- Built a cloud-based ETL pipeline (AWS Glue, S3, Athena, IAM, EventBridge Scheduler, Step Functions) to ingest, cleanse, integrate, and transform survey data from multiple sources.
- Designed aggregation and transformation logic by brand and demographic dimensions; validated outputs against expected schemas and business rules; infrastructure managed with Terraform.

Technical Skills

Programming: Python, SQL (MySQL, MS-SQL), Bash/Shell, Regular Expressions

Data Engineering: ETL Design, Data Pipeline Development, Data Cleansing, Data Integration, Data Validation, AWS Glue, Athena, S3, Step Functions, EventBridge Scheduler, Terraform, IAM, SSIS, MongoDB

Statistical & ML Modelling: Scikit-learn, Supervised Learning, Model Evaluation, Feature Importance, Gabor-Granger Pricing Models, MaxDiff, Van Westendorp, CVI/CBI Modelling, Brand Perception Modelling

Software Engineering: Git, GitHub Actions, CI/CD, pytest, Ruff, Cryptographic Data Handling

Analytics & BI: Power BI, DAX, KPI Design, Reporting, Data Modeling

Apps & APIs: Streamlit, LangChain, OpenAI API

Education

University of Ibadan

BSc. Mechanical Engineering

Ibadan, Nigeria

2016 – 2023

Turing College

Micro Masters, Data Analysis & AI Ethics

Remote, Lithuania

2024 – 2025

Udacity

Nanodegree, Data Analysis

Remote

2022

Certifications & Publications

- Microsoft Certified Power BI Data Analyst Associate (expired) – [View certificate](#)
- Co-author, *Optimization of Energy Grid Efficiency with Machine Learning* — Renewable and Sustainable Energy Reviews (Elsevier) – [View publication](#)

Additional Information

Portfolio: samietex.github.io **Languages:** English