

Nelson Peter Mwangi Mwaura

Data Engineer / Data Scientist

nmwaura132@gmail.com | linkedin.com/in/nelson-peter | portfolio-e16.pages.dev | Nairobi, Kenya

PROFESSIONAL SUMMARY

Data engineer and data scientist building production data pipelines and machine learning systems in Python and SQL. At Optiven I own the ingestion and transformation layer, moving complex datasets out of ERP systems into centralised analytics repositories with strict validation and single-source-of-truth accuracy. Outside that I designed, built, and operate Mara, a live analytics platform for the Nairobi Securities Exchange, end to end and on my own. I care most about systems that keep running once they ship: reliable ingestion, honest numbers, and pipelines that fail loudly rather than silently.

CORE COMPETENCIES

Python	SQL	ETL / ELT Pipelines	Data Warehousing	Data Quality & Validation	scikit-learn	Django	FastAPI
PostgreSQL	Docker	Power BI	Git				

EXPERIENCE

Optiven Limited

2024 to Present

Data Scientist and Data Engineer, Nairobi, Kenya

- Designed and built **ETL pipelines in Python and SQL**, ingesting complex datasets from existing ERP systems into centralised analytics repositories.
- Enforced strict validation on transformations to maintain **single-source-of-truth accuracy** across organisational data streams, with error notification rather than silent failure.
- Performed feature engineering and model development for **predictive models forecasting customer payment behaviour**.
- Worked directly in a production Django codebase to support data workflows, keeping them reusable and documented.
- Built and integrated the CRM the sales team runs on, connecting it to the existing ERP so data flows both ways, and ran the training that drove adoption.

Freelance

2022 to 2024

Data and Software Projects, Remote

- Delivered data analysis and application work for clients end to end, covering data cleaning, transformation, and delivery.
- Built integrations and data flows between systems, handling authentication and reliable data exchange.

SELECTED PROJECTS

Mara: NSE Analytics Platform (hey Mara.co.ke)

- Live analytics platform for the Nairobi Securities Exchange: ingestion, ML price forecasting, portfolio tracking, and strategy backtesting.
- Sole engineer across the stack: **Django** backend and scheduled ingestion, **PostgreSQL**, **scikit-learn** forecasting, **React** frontend, **Docker** deployment.
- Designed for unreliable upstream sources: retries, validation, and alerting on ingestion failure instead of silent data gaps.

Document Intelligence API

- FastAPI and Celery service extracting structured JSON from PDFs with GPT-4o Vision, backed by PostgreSQL and Redis.
- SHA-256 content hashing for deduplication, removing repeat model spend on identical documents, with exponential backoff on transient failures.
- Explicit job states so downstream clients render failures rather than hang.

Clinical Data Warehouse

- Architected a data warehouse centralising electronic health records from existing EHR systems.
- Built the end-to-end ETL pipeline in Python and SQL with automated validation and error notification.

Heart Disease Risk Prediction Model

- End-to-end classification workflow in Python (pandas, scikit-learn), from feature preparation through validation of model outputs.

EDUCATION

St Paul's University

May 2021 to August 2024

BSc Computer Science, major in system security

CERTIFICATIONS

Business Data Analysis and Machine Learning, DMK IT Solutions

Artificial Intelligence Analyst, IBM

Practical Network Penetration Testing, TCM Security

Hacking (and Defending) Active Directory, TCM Security

Advanced Google Analytics, Google

Android Development, Udacity

SKILLS

Data Engineering: Data pipelines, ETL/ELT orchestration, SQL, data quality and validation, structured and unstructured data processing

Data Science and ML: Python (pandas, scikit-learn), feature preparation, predictive modelling, experiment tracking with MLflow

Software Engineering: Django, FastAPI, REST API design, PostgreSQL and MySQL schema design, Docker, Git

Automation and BI: n8n workflow automation, Power BI, scheduled reporting

Languages: English (fluent), Swahili (native), Russian (intermediate)