

JESSE NGUGI GICHUKI

Data Scientist & Analyst

Nairobi, Kenya | +254 799 117179 | ngugigichuki21@gmail.com

[LinkedIn Profile](#) | [GitHub Portfolio](#) | [Dev Blog](#)

PROFILE

PROFESSIONAL SUMMARY

Results-driven **Data Scientist and Analyst** with a strong foundation in building end-to-end machine learning models, deploying structured ETL pipelines, and delivering production-ready business intelligence solution. Leveraging intensive training from a **LuxDev Data Science Bootcamp** and practical fintech/agricultural operations experience at **CoffeeKenya.Coop**, I excel at turning messy raw data into structured databases and predictive architectures. Proven ability to engineering clean datasets, manage relational databases, and design high-impact dashboards that optimize core business metrics.

CORE COMPETENCIES

Data Science & Machine Learning: Predictive Modeling, Feature Engineering, Classification/Regression, Train-Test Splitting, Class Imbalance Handling (Scikit-Learn), Model Evaluation, Automation.

Data Engineering & ETL: Pipeline Automation, Workflow Scheduling (Apache Airflow), Data Cleaning (Pandas, NumPy), Database Management (PostgreSQL, DBeaver), Data Ingestion.

Analytics & Business Intelligence: Advanced Reporting, KPI Tracking, Unit Economics, Interactive Dashboards (Power BI, Microsoft Excel, Power Query, DAX), Data Modeling, Schema Architecture.

TECHNICAL TOOLSTACK

Programming Languages: Python, Java, JavaScript, SQL (PostgreSQL).

Libraries & Frameworks: Pandas, NumPy, Scikit-Learn, SciPy, Matplotlib, Seaborn, Anaconda.

Tools & Platforms: Apache Airflow, Power BI, MS Excel, DBeaver, Git/GitHub, AWS, Azure, Google Cloud, DigitalOcean.

PROFESSIONAL EXPERIENCE

Data Analyst / FP&A Resource | CoffeeKenya.Coop (Nairobi, Kenya)

2024 – Present

ETL & Automation: Built automated Python scripts to clean messy transactional and member datasets, handle missing values, and standardize variable scales, reducing manual prep work by 35%.

Dashboard Development: Engineered dynamic Power BI portfolios and multi-entity Excel dashboards leveraging Power Query and DAX to trace producer delivery metrics, payments, and unit economics.

Database Administration: Structured and managed relational tables using PostgreSQL and DBeaver to store cleaned agricultural data, optimizing complex query execution times.

Business Intelligence: Designed board-ready reporting packs on a fixed monthly cadence, interpreting variance analysis and data trends directly for cooperative leadership stakeholders.

KEY DATA PROJECTS

Churn Predictor (End-to-End Machine Learning) | [GitHub Link](#)

Developed a complete machine learning workflow to predict customer attrition using Python and Scikit-Learn.

Conducted deep exploratory data analysis (EDA) using Seaborn and Matplotlib to visually isolate categorical drivers of churn like contract structures and service types.

Engineered clean numeric pipelines using `pd.to_numeric(errors='coerce')` and missing value imputation via `.fillna()` to convert loose text fields into model-ready features.

Executed stratified train-test splits (`test_size=0.20`, `stratify=y`) to maintain class balance; trained, saved, and structured standalone scripts (`train.py`, `features.py`, `predict.py`) alongside an operational application API directory.

Sales Dashboard ETL & Workflow Automation | [GitHub Link](#)

Designed a programmatic data pipeline to ingest, clean, and structure raw daily sales revenue fields into an analytical database.

Automated script scheduling using Apache Airflow Directed Acyclic Graphs (DAGs) to monitor workflow tasks sequentially and prevent upstream execution failures.

EDUCATION & CERTIFICATIONS

Professional Bootcamp Certification in Data Science & Analytics | LuxDevHQ (2025)

Comprehensive Data Modeling & Visual Analytics Mastery | Microsoft Power BI / Excel Focus (2024)