

Samwel Maroa Waiyaki

Data Scientist | Engineer | AWS Certified Solutions Architect | Business Analyst

Contact

Phone: +254713645965

Email: Samwelmaroawaiyaki01@gmail.com

LinkedIn: [Click here](#)

Github: [Public Repositories](#)

Location: Nairobi, Kenya

Education

Strathmore University - Kenya

May 2024 – June 2026 (Expected graduation)

MSc. Data Science and Analytics

Explore Data Science Academy

Jan 2022 – Oct 2022

Data Science Professional Certificate

Multimedia University of Kenya

2015-2019

BSc. Mathematics and Computer Science
(Statistics Major)

Second-Class Honours (GPA 3.6)

HSE University (Coursera)

Natural Language Processing

Skills

Data analysis & Machine learning

Python, Pandas, NumPy, Scikit-learn, TensorFlow, Keras, Huggingface Transformers, Apache Kafka, Apache Spark, Dagster, R, RStudio,

Web Application Frameworks

Flask, Django, Streamlit, R-Shiny, HTML, CSS, JavaScript

Professional Summary

I am a Data Scientist/Engineer with 5 years of experience in designing and implementing projects. I excel at gathering, cleaning, and organizing data for both technical and non-technical stakeholders. I am skilled in building machine learning models and APIs for deploying machine learning products. Additionally, I have experience in transfer learning, leveraging state-of-the-art pre-trained transformers on custom datasets to produce high-quality, accurate models. My proficiency in Apache technologies such as Kafka, Spark, and Dagster enables me to construct robust ETL and machine learning pipelines for production environments. I am organized, motivated, and diligent, with a strong background in Data Analysis, Machine Learning, Business Intelligence, and programming. I possess advanced proficiency in Python, R, SPSS, MS Excel, Power BI, Tableau, and SQL.

Experience

Telkom Kenya, Technical Business Analyst, Sep 2022 – Current

- I analyzed and created a credit scoring ML model for the hustler-fund mobile lending platform sponsored by government of Kenya for Telkom Kenya mobile subscribers.
- I built and maintaining machine learning models that helps in predicting customers who are likely to churn so as to take necessary measures for retention. This has helped the department improve retention by 25%
- I built and maintain voice traffic ETL pipeline using Python, Oracle SQL and Apache Dagster for orchestration.
- Building time-series forecasting models that have helped in account managers in projecting their revenue gains on month-on-month basis and therefore enhanced their drive to be on target.
- Building and maintaining BI dashboards for the department.
- Conducting A/B test analysis to discover viability of new products and strategies that support wholesale business for broadband and connectivity.
- Maintain database holding voice traffic data for wholesale department.
- Ensuring efficient, timely and accurate reporting through automation of reports

StepWise Inc, Data Scientist/BI Specialist, Feb 2021 – August 2022

- Building, optimizing, and maintenance of machine learning models for the business
- Building time-series and forecasting models for the company's various business aspects
- Developing BI reports and dashboards for executive leadership
- Performing general data analysis to identify trends, and patterns, and make inferences from business data
- Building recommendation reports guided by insights from data from strategic decision-making.
- Using business data to investigate efficiency levels in the business processes and suggest optimization mechanisms

Zalego Academy, StepWise Inc. Zalego academy is a subsidiary company of StepWise Inc.

Business Intelligence

Power BI, Tableau, Excel

Cloud Platform and Databases

Aws Cloud, Oracle SQL, MySQL, PostgreSQL

Accomplishments

1. Assisted to create a Credit Scoring model for government contracted Hustler-Fund Wallet
2. AWS Certified Solutions Architect Associate
3. Deployed 6 data solutions as web-applications
4. Deployed 1 Voice traffic ETL pipeline to Production

References

Francis Yego:

Senior Data Analyst – Alliance Bioversity & CIAT

Phone: +254 707 825181

Email: F.Yego@cgiar.org

Brian Aholi:

Lead Data Scientist – Diamond Trust Bank Kenya Limited.

Phone: +254 719 592496

Email: blipesa@dtbafrica.com

Elijah Abuta

Position: Business Analyst

Company: Saliency Consulting LLC

Phone: +254 727 966444Email:

elijah.motari@saliencyconsulting.ae

- Trained 3 cohorts of data analysis students and achieved over 80% pass rate:
- Made logistical planning accurate by building a student intake forecasting model.
- Organized and delivered workshops on data analysis to the corporate customers
- Facilitated data science webinars in the academy
- I developed data analysis content for the academy
- Performed general data analysis in the academy for insight discovery

Freelance, Data Scientist. Apr 2019 - Current

- Developing Machine learning-based web application (Flask, Django)
- Building data APIs
- Deep learning and Machine learning modeling
- Building product recommender systems.
- Natural language processing
- Time-series forecasting models
- Statistical analysis and modeling
- Power BI and Tableau BI reports and dashboards

Some of My Key Data Projects

1. **Hustler Fund Credit Scoring Model:** This project was funded by the government of Kenya with the aim of establishing a lending platform to provide citizens with access to credit at affordable interest rates. As part of this project, we utilized data from Telkom Kenya customers, including mobile money transfers, airtime usage, and internet data consumption. My role involved developing a credit scoring model that calculates customers' credit limits. These limits are then adjusted based on customers' repayment behavior over time.
2. **Voice ETL Pipeline:** In this pipeline project, I designed and implemented a system to fetch real-time Call Detail Records (CDRs) from a switch database and pricing data from a dedicated price list database. The pipeline's transformation phase involved applying the appropriate rates to each CDR based on the destination of the call, computing the call charges based on its duration and the applicable rate. This process ensured that no calls were duplicated, and the transformed data was seamlessly loaded into the billing database. Technologies utilized for this project included Python for scripting, SQL for database operations, and Dagster for orchestrating the end-to-end ETL workflow.
3. **Sentiment Analysis using Transfer Learning:** In this personal project, I utilized the Bert Transformer model for performing sentiment analysis through transfer learning. Specifically, I fine-tuned the model using the Olist E-commerce customer review dataset. Despite encountering significant imbalance in the target variable, I successfully developed a model achieving a 93% F1-score. Such high performance would typically pose challenges when attempting to build machine learning models from scratch.
4. **ETL Pipeline for Olist E-commerce dataset using Python and Dagster.**
5. **GitHub:** <https://github.com/Maroa-Waiyaki?tab=repositories>