

**MUTUA SIMON MUSYOKI**  
**DATA ANALYST | DATA SCIENTIST | STATISTICIAN**

+254 792 162 750 | [simonmusyoki2019@gmail.com](mailto:simonmusyoki2019@gmail.com) | [linkedin: simon-musyoki](#) | [Github - Simonrank](#) | Kenya

---

## SUMMARY

Data Analyst with hands-on experience in SQL, Python, Power BI, and Tableau, specializing in transforming raw data into clear, actionable insights for decision-making. Strong background in applied statistics, dashboard automation, KPI development, data cleaning, and exploratory analysis. Experienced working on healthcare, community development, and operational data projects, including predictive modeling and medical information retrieval solutions. Passionate about using data and AI to improve healthcare outcomes and operational efficiency.

## SKILLS

### Technical & Analytical Skills

- SQL
- Power BI, Tableau
- Data modeling & KPI development
- Data visualization & reporting
- Python
- Data cleaning & wrangling
- Machine learning
- Basic cloud data concepts

### Soft Skills & Professional Skills

- Problem-solving
- Collaboration
- Project coordination
- Critical thinking
- Communication

## Projects

### [Heart Failure Prediction \(June-2025\) - https://github.com/Simonrank/Heart\\_Failure\\_Prediction](#)

- Addressed the problem of early detection of heart failure risk, which affected patient outcomes, care prioritization, and hospital readmission rates.
- Combined 900+ patient records to engineer clinical features such as age, cholesterol, and blood sugar for predictive modeling.
- Built classification models using Python (Pandas, Scikit-learn) with cross-validation and ROC-AUC evaluation.
- Delivered an ROC-AUC of 0.92, enabling hospitals to identify high-risk patients faster, optimize care prioritization, and reduce readmission rates, demonstrating the impact of BI and machine learning on operational efficiency and patient outcomes

### [Medical Chatbot for Cancer Information Retrieval chatbot \(July 2025\) - Collaboration-](#)

#### [https://github.com/joyaran/Cancer\\_NLP\\_project](https://github.com/joyaran/Cancer_NLP_project)

- Addressed the problem of cancer-related misinformation and limited access to reliable medical answers for patients and the public.
- Processed the MedQuAD dataset (729 XML Q&A pairs) using a custom text preprocessing pipeline.
- Applied TF-IDF and BioBERT embeddings to build a hybrid information-retrieval system.
- Evaluated similarity models and optimized retrieval confidence thresholds.
- Achieved 85.71% retrieval accuracy and deployed a Gradio-based chatbot for fast, trustworthy cancer information access.

### [YAKWETU™ Climate-Based Variable Pricing Project \(August-2025\) – Collaboration](#)

#### [https://github.com/joyaran/Yakwetu\\_project](https://github.com/joyaran/Yakwetu_project)

- Addressed the problem of unpredictable movie demand influenced by changing weather patterns, which affected pricing, scheduling, and revenue optimization.
- Combined 30K+ user activity records with 9K+ weather data points to analyze audience demand patterns.
- Built Power BI dashboards and forecasting models using XGBoost and Prophet.
- Delivered 77.9% accurate demand forecasting and insights projecting up to 11.5% revenue increase through dynamic pricing strategies.

# EXPERIENCE

---

## Attachment, Kenya Bureau of Standards National Quality Institute( NQI )Department April - July 2024 | Nairobi

- Developed interactive dashboards in Power BI and Tableau, reducing reporting time by 25% and enabling faster data-driven decision-making for stakeholders.
- Coordinated logistics and managed the National Quality Institute (NQI) Integrated Management Systems (IMS), ensuring 100% compliance with quality standards.
- Collaborated on module development and conferencing initiatives, strengthening stakeholder engagement and expanding NQI membership reach.

## Survey Analyst, Reuben Centre September 2024 | Nairobi

- Led a comprehensive survey engaging 300+ community members, staff, and partners to identify key organizational strengths and value propositions.
- Analyzed and synthesized complex survey data, uncovering critical opportunities and delivering actionable recommendations that directly guided the Centre's rebranding strategy
- Translated business needs into structured survey questions, ensuring data collection aligned with strategic objectives.
- Presented data findings and strategic insights to leadership, influencing key decisions for future initiatives.

## Enumerator, Ministry of Labour and Social Protection October 2024 | Nairobi

- Conducted household surveys and in-depth interviews, ensuring 100% data completeness and accuracy for welfare and social safety net programs.
- Guided diverse respondents through complex questionnaires, improving response clarity and reducing survey errors by 15% through effective communication.
- Collaborated effectively with field teams to streamline data collection processes and ensure timely submissions.

## Volunteer Intern | Salama Youth Hub October 2024- | Nairobi

- Analyzed beneficiary data to identify program effectiveness and areas for improvement, directly informing monthly program strategy updates.
- Developed monthly reports that informed program strategy, improved service delivery, and strengthened stakeholder engagement.
- Facilitated weekly support groups and open discussions, increasing youth participation in programs by 30%.
- Contributed to data-driven decision-making by tracking key metrics and presenting findings to the leadership team.

# EDUCATION

---

- Moringa School | Full Advanced Data Science course

Graduated November 14,2025

- Karatina University| Bachelor of science in Applied Statistics with Computing

Second Class Honors Lower Division

- Allison| Certificate in Google Analytics 4

- Miccato America Share | certificate in computer packages

- Makindu Boys' School | Kenya Certificate of Secondary School (KCSE)

- Volunteer Experience-Volunteered at Moringa school career fair TecheXpo 2025