MATHEW SHEM

BIOSTATISTICIAN | DATA SCIENTIST | DATA ANALYST | AI BUILDER

PERSONAL INFORMATION

Full Name: Mathew Shem
Email: mailto:mathewshem@yahoo.com
Phone: +254768944928
Postal Address: 62000 – 00200, Nairobi – Kenya
LinkedIn: Mathew Shem
Website: Mathew Shem
GitHub: Mathew Shem
Portfolio: Mathew Shem

PROFESSIONAL SUMMARY

I am a professional Biostatistician with a Bachelors' degree from Jomo Kenyatta University of Agriculture and Technology(JKUAT), Faculty of Mathematics and Physical Sciences and the Department of Statistics and Actuarial Sciences. I have hands-on experience in Clinical Trials, Genomics, Survival Analysis, Spatial Epidemiology, Statistical Modeling, and Public Health Data Analysis. My final year research applied Generalized Linear Modelling (GLM) and Geographically Weighted Logistic Regression (GWLR) to model infant mortality in Kenya using the 2022 Kenya Demographic and Health Survey (KDHS) dataset - incorporating spatial diagnostics including Local Indicators of Spatial Association (LISA) and Moran's I to detect geographic clustering and spatial autocorrelation in health outcomes. My final year project was awarded 'Best in the Department' and I got a chance to present in the 2nd Conference of Mathematics and Applied Sciences.

Beyond my academic work, I have applied geospatial methods in practice: at KEMRI's Alupe facility I conducted spatial disease surveillance using R packages including sp, rgdal, and tmap, and I collaborated with Hamidu Health Center in Ghana on a QGIS-based project mapping community distances to medical facilities to support health access planning.

I have further contributed to public health analytics through participation in an Africa CDC COVID-19 digital communication study, Kenyatta National Hospital(KNH)'s Medical Research Department and an infectious disease predictive modeling hackathon at JKUAT (HERMIS). My technical toolkit spans R, Python, SQL, SPSS, STATA, QGIS, and REDCap, with advanced additional experience in data visualization using Power BI and Tableau. I am deeply motivated by the role of analytics in reducing disease burden and strengthening health systems, and I am eager to apply these skills to deliver measurable, data-driven impact for organizations and clients I work with.

PROFESSIONAL EXPERIENCE

Data Analyst

MobileMart Services Limited – EABL Distributor
FMCG Industry | August 2026 – Present

- Analyze sales, route-to-market, outlet, and distribution data to evaluate business performance and identify trends, gaps, and growth opportunities.
- Develop and maintain analytical reports and interactive dashboards using Microsoft Excel, Power BI, and Tableau to support sales and operational decision-making.
- Analyze route sales performance, outlet purchasing patterns, product movement, and sales achievement against targets and previous periods.
- Generate actionable business insights for sales and management teams, highlighting underperforming areas, product gaps, outlet opportunities, and performance trends.
- Perform data extraction, cleaning, transformation, validation, and analysis from distributor management and financial systems.
- Prepare periodic sales and performance reports, including current-versus-prior-year analysis, target achievement, outlet performance, and product/category analysis.
- Support management with data-driven analysis for sales planning, performance monitoring, and operational decision-making.
- Collaborate with sales, accounts, administration, fuel, and other business functions to identify reporting and analytics requirements.
- Developing DATANOVA, an automated analytics and reporting system designed to automate departmental reporting across Sales, Fuel, Accounts, Administration, and Data functions.
- Designing DATANOVA as a scalable distribution analytics SaaS platform that can potentially be adopted by FMCG distributors to automate reporting, analytics, and business intelligence workflows.
- Apply data analytics, business intelligence, and automation principles to improve reporting efficiency and reduce manual data-processing workflows.

Lead Tutor & Database Administrator

StatQuestJourney Hub
January 2023 – December 2025

- StatQuestJourney Hub is a training and consultancy platform focused on data science, statistics, analytics, and research support across East and West Africa.

Key Responsibilities:

- Designed and delivered more than ten full training programs in Data Science, Data Analysis, Programming, and Statistics covering R, Python, Power BI, Tableau, SQL, and Excel.
- Developed and maintained databases for student records, course materials, and analytics resources while ensuring proper data management and integrity.
- Authored educational content including blogs, analytical guides, and statistical tutorials covering machine learning models, data analysis workflows, and statistical methodologies.
- Organized multiple hackathons, webinars, and analytics competitions aimed at building practical data science skills among students and early-career professionals.
- Provided consultancy services involving statistical modeling, predictive analytics, and research support for academic and industry clients.
- Mentored learners and researchers across Kenya, Nigeria, and Uganda in areas including data cleaning, visualization, statistical inference, and machine learning.
- Head of general IT related issues and domain management using Namecheap, Namecheap's cpanel, Host Africa, WIX and Wordpress

Project Researcher & Webinar Presenter

Africa CDC COVID-19 Communication Study
July 2024 – February 2025

- Participated in a research initiative analyzing digital communication trends during the COVID-19 pandemic across African countries.
- Responsibilities:
- Collected and scraped social media data from Facebook and Instagram using Python libraries including BeautifulSoup4, Instaloader, FacebookScraper, and Pandas.
- Performed data cleaning, preprocessing, and exploratory data analysis using R and Python.
- Analyzed patterns in COVID-19 communication strategies, identifying trends in public health messaging and engagement across multiple African regions.
- Conducted statistical analysis and trend evaluation using R Studio.
- Presented research findings in an international webinar hosted by StatQuestJourney Hub attended by participants from Kenya, Nigeria, and Uganda.

- Delivered technical sessions demonstrating real-world data collection, preprocessing, visualization, and research analysis workflows.

Hackathon Competitor – Winner

Gomycode Tableau Hackathon
December 2024

- Developed a professional business intelligence dashboard analyzing trends in the Electric Vehicle industry.
- Key Contributions:
- Built an advanced interactive Tableau dashboard analyzing market trends, product adoption patterns, and regional demand.
- Applied calculated fields, filters, dashboard actions, and storytelling techniques to present analytical insights effectively.
- Won the hackathon competition held in Nairobi, Upper Hill.

Predictive Modeling Contributor

JKUAT Infectious Disease Modeling Hackathon
August 2024

- Collaborated in a data science team developing predictive models to analyze infectious disease outbreak patterns.
- Responsibilities:
- Performed data cleaning and preprocessing using Excel and R.
- Conducted advanced data manipulation using R libraries including dplyr and tidyr.
- Developed predictive models using statistical and machine learning techniques.
- Contributed to analytical workflows for disease prediction and health analytics.

Freelance Data Consultant

- Independent Projects (Ongoing)
- Provide statistical and analytics consulting services to academic researchers, students, and business clients.

- Services include:
- Statistical analysis using R, SPSS, STATA, and Excel.
- Development of Power BI and Tableau dashboards for business intelligence and reporting.
- Support for thesis research including hypothesis testing, regression modeling, and survival analysis.
- Machine learning model development and predictive analytics.
- Consultation in clinical trials, public health data analysis, and research methodologies.
- Conducted statistical tests including Chi-square, regression analysis, and multivariate statistical modeling.

Volunteer Data Mentor

DataCamp (Everything Data Program)
January – April 2024

- Mentored learners in SQL, Excel, and Power BI during live online sessions.
- Responsibilities:
- Guided learners through hands-on exercises involving real datasets.
- Explained complex analytical concepts in accessible formats.
- Provided live support and feedback during data analysis assignments.
- Facilitated practical learning sessions focused on business intelligence and data analysis workflows.

Attachee – Biostatistics

Kenya Medical Research Institute (KEMRI), Alupe – Busia
June 2024

- Worked on spatial epidemiological data analysis supporting disease surveillance initiatives.
- Responsibilities:
- Applied spatial statistical techniques to analyze geographic patterns of disease incidence.
- Used R packages including sp, rgdal, and tmap to visualize spatial distributions.

- Modeled spatial autocorrelation using Moran's I and variograms.
- Integrated geographic information systems with epidemiological data to support disease control planning.

Attachee – Medical Research Department

Kenyatta National Hospital (KNH)
May – August 2025

- Worked with clinical research teams supporting healthcare analytics and patient outcome studies.
- Responsibilities:
- Conducted exploratory data analysis, data cleaning, and statistical modeling using R Studio.
- Implemented regression analysis, hypothesis testing, and survival analysis for clinical datasets.
- Developed reproducible analytical scripts for patient outcome tracking and longitudinal analysis.
- Designed and managed data collection instruments using REDCap.
- Ensured data quality through validation checks and logical controls in clinical databases.
- Compiled and aggregated data using Excel to generate weekly and monthly reporting dashboards.
- Developed pivot tables, lookup functions, and charts to visualize healthcare metrics and research indicators.

ARTIFICIAL INTELLIGENCE & INNOVATION PROJECTS

EduNav+ – AI Assistive Navigation System for the Blind

- STEM Hackathon for Persons with Disabilities
- Developed an assistive artificial intelligence system designed to help visually impaired students navigate academic environments independently.

Key Features:

- Real-time object detection and obstacle avoidance using computer vision techniques.
- Live optical character recognition (OCR) system capable of reading classroom board content and printed materials.
- Multilingual translation capabilities integrated with Braille-compatible outputs for accessible learning.
- AI-driven navigation guidance enabling blind students to locate classrooms while avoiding obstacles.
- This project demonstrated applications of:
- Computer Vision
  Assistive Artificial Intelligence
  Edge AI Systems
  Human-Centered Technology Design

LARGE LANGUAGE MODELS (LLMs), AI EVALUATION & MACHINE LEARNING SYSTEMS

Developed and evaluated machine learning and AI-powered solutions using Python, Scikit-Learn, TensorFlow, PyTorch, and modern AI development frameworks. Experienced in designing benchmark tasks, creating evaluation rubrics, assessing model outputs, and validating AI-generated content across technical and analytical domains. Applied natural language processing (NLP), predictive modeling, computer vision, and statistical learning techniques to solve real-world problems. Familiar with Large Language Models (LLMs), prompt engineering, Retrieval-Augmented Generation (RAG), AI-assisted development workflows, model performance evaluation, and systematic testing of AI systems for accuracy, reliability, and business applicability. Experienced in translating complex technical requirements into scalable data products, analytical solutions, and production-ready machine learning workflows.

- Evaluated AI-generated analytical reports, code outputs, dashboards, and technical documentation for accuracy and usability.
- Designed machine learning workflows involving data preprocessing, feature engineering, model training, validation, and performance optimization.
- Applied NLP techniques including text preprocessing, sentiment analysis, document classification, and information extraction.
- Built AI-powered applications leveraging modern LLMs and AI-assisted development tools.
- Developed predictive models using regression, classification, ensemble learning, and statistical methods.

- Conducted model evaluation using accuracy, precision, recall, F1-score, ROC-AUC, and cross-validation techniques.
- Utilized Git, GitHub, Jupyter Notebooks, and collaborative development workflows for AI and data science projects.

LARGE LANGUAGE MODELS (LLMs), AI EVALUATION & MACHINE LEARNING SYSTEMS

- Developed robust social media data collection pipelines using Python (BeautifulSoup4, Instaloader, FacebookScraper) to scrape and analyze digital communication trends during the Africa CDC COVID-19 study.
- Built automated data ingestion, cleaning, and trend analysis workflows for large-scale public platforms, identifying engagement patterns and momentum signals across multiple African countries.
- Created interactive dashboards (Tableau & Power BI) to visualize growth metrics, audience trends, and key insights — adaptable to music/creator economy analytics (TikTok, Instagram, YouTube, Spotify signals).
- Experienced in ethical data practices, API integration, and designing compliant automation workflows for public data sources.

BUSINESS INTELLIGENCE & DATA VISUALIZATION

Tableau

- Winner of the Gomycode Tableau Hackathon for developing an advanced dashboard analyzing the Electric Vehicle industry.
- Attended Salesforce Trailblazers Tableau training program involving intensive hands-on sessions in advanced data visualization and dashboard design.
- Built professional dashboards integrating multiple data sources including Excel datasets and structured databases.
- Applied advanced visualization features including calculated fields, parameters, dashboard actions, and visual storytelling techniques.
- Delivered workshops teaching Tableau dashboard design and data storytelling practices.

Power BI

- Developed enterprise-level dashboards using Power BI for real-world business datasets.

- Integrated multiple data sources including SQL databases and Excel datasets.
- Applied Power Query for advanced ETL processes and data transformation.
- Developed KPI monitoring dashboards and visual reporting systems.
- Taught Power BI courses covering dashboard development, DAX calculations, and data storytelling.

PROGRAMMING & DATA SCIENCE EXPERTISE

R Programming

- Developed statistical models using packages including caret, ggplot2, and tidyverse.
- Conducted advanced statistical testing, hypothesis testing, and time-series forecasting.
- Developed interactive dashboards using Shiny.
- Created animated visualizations using ggplot and gganimate.
- Performed web scraping using rvest and httr libraries.
- Conducted post-hoc statistical tests and advanced statistical modeling.

Python Programming

- Used BeautifulSoup4 for large-scale social media data scraping.
- Developed machine learning models using scikit-learn.
- Performed exploratory data analysis using pandas and matplotlib.
- Automated analytics reporting pipelines using Jupyter notebooks.
- Built APIs supporting predictive analytics systems.
- Taught Python programming fundamentals for data science applications.

SQL

- Developed structured query systems and relational database schemas.
- Performed advanced queries including joins and stored procedures.

- Optimized queries for analytics dashboards and reporting systems.
- Integrated SQL outputs with Power BI and Tableau.
- Performed database creation and interaction using Python libraries including pymysql.
- Conducted SQL-based data cleaning and transformation workflows.

STATISTICAL SOFTWARE & ANALYTICS TOOLS

- SPSS – Regression modeling, hypothesis testing, and survey data analysis.
- STATA – Panel data analysis, regression modeling, and health economics studies.
- SAS – Large-scale survey data analysis, macro programming, and statistical modeling.
- Excel – Advanced dashboards, forecasting tools, pivot tables, and business reporting.
- Jamovi – GUI-based statistical analysis for teaching and quick statistical testing.
- NVivo – Qualitative research coding and mixed-methods data analysis.

DATA COLLECTION & PUBLIC HEALTH SYSTEMS

- CommCare
  ODK
  REDCap
  KoboCollect
- Used for mobile data collection, health research surveys, and public health consultancy projects.

ACADEMIC HISTORY

- 2022 – 2026
  Bachelor of Science in Biostatistics
  Jomo Kenyatta University of Agriculture and Technology (JKUAT)
- 2018 – 2021
  Kenya Certificate of Secondary Education
  Grade: B+

- 2007 – 2017
  Kenya Certificate of Primary Education
  Grade: A

BLOGS, VIDEOS & PUBLICATIONS

- Author of educational blog series covering machine learning models and applied statistics.
- Published a comprehensive PDF guide on statistical tests used in research.
- Create educational YouTube tutorials covering R programming, Excel analytics, and Python programming.
- Authored technical articles covering R, Python, Julia, Data Science, Data Analysis, and applied statistical techniques.

LEADERSHIP & EXTRACURRICULAR ACTIVITIES

Chairperson
JKUAT Biostatistics Students Association (JKUBSA)

- Facilitated collaboration between students and research institutions including KEMRI, ICIPE, ILRI, KNH, AMREC, and KALRO.
- Organized research networking events and professional development workshops.

Event Organizer – JKUBSA

- Hosted multiple academic events including networking sessions with Kisii University Statistics Club.

Student Leadership Representative

- Served as Non-Residential Representative responsible for accommodation, security, catering, and bursary committee participation within the university leadership structure and also as an international students representative in the congress
- Community outreach volunteer organizing professional meetups and educational workshops.

MATHEW SHEM

BIOSTATISTICIAN | DATA SCIENTIST | DATA ANALYST | AI BUILDER

KEY SKILLS & QUALIFICATIONS

- Advanced Statistical Modeling and Hypothesis Testing
  Data Cleaning, Transformation, and Visualization
  Predictive Modeling and Machine Learning
  Public Health and Epidemiological Data Analysis
  Artificial Intelligence and Computer Vision Applications
  Business Intelligence and Data Visualization
  Research Methodology and Scientific Writing
  Curriculum Development and Data Science Education

Technical Expertise:

- R
  Python
  SQL
  SPSS
  STATA
  SAS
  Power BI
  Tableau
  Excel
  MATLAB

Additional Skills:

- Graphic Design (Canva, Adobe Illustrator, CapCut)
- Systems and web development using MERN, PHP, Clerk, Cloudinary
- IT Support i.e. DNS configuration, Email management, Database administrator, Automation setups, Namecheap expert, Wix & Wordpress expert

LANGUAGES

English – Fluent
Swahili – Fluent
Pidgin English

REFEREES

1. **Dr. Henry Kissinger**
   Senior Lecturer – Department of Statistics and Actuarial Sciences, JKUAT

P.O. Box 62000-00200, Nairobi Kenya
Phone: +254723440110
Email: henryathiany@gmail.com

2. **Dr. Caroline Mugo**
   Senior Lecturer – Department of Statistics and Actuarial Sciences, JKUAT
   P.O. Box 62000-00200, Nairobi Kenya
   Phone: +254733388887
   Email: cmugo@jkuat.ac.ke

3. **Mohamed Issa**
   Senior Lecturer – Department of Botany, JKUAT
   P.O. Box 62000-00200, Nairobi Kenya
   Phone: +254732511644
   Email: mohamedissa@jkuat.ac.ke