MICHAEL NSIAH-NIMO

DATA ANALYTICS PROFESSIONAL – DATA ANALYSIS & DATA SCIENCE

• Website • Lincoln, Nebraska • +1 (915) 246-5125 • michael20129648@gmail.com

PROFESSIONAL SUMMARY

Research Statistician and Certified DevOps Engineer with 7+ years of experience in advanced analytics, machine learning, and cloud-native solutions. Skilled in developing predictive models, scalable AI/ML pipelines, and automated workflows with AWS. Proficient in statistical modeling, supervised and unsupervised learning, and data engineering for large-scale, high-dimensional datasets. Experienced in data integration, standardization, and reproducible analytics from diverse sources to ensure interoperability and quality. Contributed to 12+ federally funded projects, delivering actionable insights, workflow automation, and decision-support tools that improved efficiency and empowered cross-functional teams

Education

PhD: Statistics, Expected 2029, **University of Nebraska Lincoln**
MSc: Statistics, 2017, **University of Texas - El Paso**
BSc: Actuarial Science, 2014, **Kwame Nkrumah University of Science and Technology**

SKILLS

- **Visualization & Reporting:** Tableau, Power BI, R Shiny, RMarkdown, Quarto
- **Programming & Statistical Modeling:** R (lme4, tidyverse), Python (pandas, scikit-learn, XGBoost), SAS, SQL, Bash, PowerShell
- **Databases & Cloud Platforms:** AWS (SageMaker, Lambda, S3), Google BigQuery, Oracle, SQL Server, MySQL, Netezza, Snowflake,
- **ML/Stats**: Scikit-learn, TensorFlow, PyTorch, glmnet, XGBoost, SHAP, Bayesian models
- **Data Engineering & Automation:** ETL Scripting (YAML, JSON, Git Bash), SQL Data Pipelines, Cloud Data Engineering, Workflow Automation
- **Advanced Analytics & Methods:** Longitudinal & Mixed-Effects Modeling, Predictive Analytics (GLMs, Penalized Regression, DNNs), Causal Inference, Survival Analysis, NLP, Experimental Design, Survey Design & Analysis
- **Systems & Administration:** Linux/Windows Admin, Active Directory, DNS, DHCP, VMware
- **Communication & Reporting:** Technical Documentation, Policy Briefs, Stakeholder Presentations

CERTIFICATIONS

- **Advanced Learning Algorithms - Stanford University**
- **AWS Certified Solutions Architect – Associate**

WORK HISTORY

Research Statistician, 01/2020 to Present
Border Biomedical Research Center (BBRC), University of Texas at EL Paso – El Paso, TX

- Led institutional and program analytics, designing KPI frameworks for 12+ NIH sub-projects and producing technical reports and policy briefs that sustained >$2M in extramural funding.
- Led modeling and experimentation using real-world data (EHR, insurance cost, clinical procedures) to uncover cost drivers and inform reimbursement decisions in health-related financial analytics.
- Built and maintained interactive dashboards and real-time visualizations in Tableau and R Shiny and Quarto.
- Developed and validated ML predictive models (robust regression, penalized regression, deep neural networks) and applied unsupervised methods (clustering, PCA) to uncover risk factors, patterns, and performance drivers.
- Engineered cloud-native analytics workflows with AWS SageMaker, Lambda, and Step Functions, and deployed applications using Docker, Kubernetes and GitHub Actions for CI/CD.
- Designed and executed analysis plans for experimental and longitudinal studies, incorporating, stratification, and power calculations and multiplicity adjustments of error rates using R and SAS.
- Conducted quality control and validation of analytical outputs (tables, visualizations, reports), created data validation specifications (DVS), and ensured compliance with data standards for regulatory and business reporting.
- Extract and analyze unstructured institutional and clinical text data, generating insights for research and policy evaluation with transformer-based LLMs

FEATURED PROJECT: Risk Index Development
NIH-Funded Project | Border Biomedical Research Center, University of Texas at El Paso
Tools: R, Python (pandas, scikit-learn), XGBoost, GLMs, PCA, Tableau, R Shiny
Context: Chronic low-grade inflammation is a key predictor of aging-related diseases and health disparities. The goal was to develop a **composite SCI index** from high-dimensional cytokine biomarker data to quantify inflammation burden across populations and assess disparities in Hispanic-Origin Individuals (HOI).
Action:

- Collected and processed **50+ cytokine biomarkers** from Luminex assays across multiple NIH-funded studies.Applied **dimension reduction techniques (PCA, variable selection, penalized regression)** to identify the most informative cytokines. Built an **SCI score** using GLMs and gradient boosting (XGBoost) to track inflammation levels and predict aging-related health outcomes. Developed **interactive visualizations (Tableau, R Shiny)** for clinicians and researchers to monitor SCI trajectories

Results:

- Identified **key cytokines** driving variability in inflammation across the cohort. The SCI score explained **>40% of variance** in age-related disease progression markers. Findings supported a **$1M NIH grant renewal**, positioning SCI as a potential **biological age metric** in health disparities research.

Growth / Next Steps:

- Expanded modeling to incorporate **EHR and lifestyle data** for a multi-domain inflammation index.Proposed using the SCI framework for **longitudinal tracking of at-risk populations**, enabling targeted interventions in public health and workforce well-being programs.

Data Analyst – CoS Research, 07/2018 to 07/2020
University of Texas – El Paso, TX

- Designed Tableau dashboards tracking enrollment, graduation, and academic program performance for institutional research reporting.
- Delivered executive-ready reports summarizing enrollment trends and faculty productivity for academic leadership and policy stakeholders.
- Extracted and prepared multi-source patient data from NIH All of Us and PHIX Health Information Exchange, integrating EHR and claims data covering visits, labs, imaging, diagnoses, and payments.
- Designed interactive dashboards for clinicians and administrators to visualize high-cost drivers and equity gaps in access and reimbursement.

Cloud DevOps Engineer, 03/2022 to 07/2025
Technology Excellence Services – Remote (Freelance)

- Supported data platform modernization with AWS infrastructure, ensuring secure and scalable environments for analytics workloads.
- Automated ETL workflows and data pipeline deployments using GitHub Actions and CloudFormation, reducing manual reporting delays.
- Developed Ansible playbooks in YAML to enforce configuration management and maintain consistent server provisioning across environments.

SELECT PORTFOLIO PROJECTS

Quantifying Sales Uplift with Causal Impact Analysis - 🔗 **View project**

- **Context**: An organization needed to evaluate whether a new outreach campaign led to measurable improvement in engagement metrics, accounting for seasonal and external factors.
- **Action**: Collected time-series data before and after the campaign launch, integrating engagement, demographic, and regional variables. Applied Bayesian structural time-series modeling (CausalImpact) to estimate the counterfactual—what outcomes would have been without the campaign. Validated model assumptions and visualized results in R Shiny dashboards for decision-makers.
- **Result**: Isolated the campaign's true causal effect, showing a statistically significant 12% increase in post-campaign engagement after controlling for confounding factors. Findings guided resource allocation decisions for future outreach efforts.
- **Growth / Next Steps**: Recommended longitudinal expansion of the model to evaluate the cumulative impact of multiple campaigns over time. Proposed building automated dashboards for real-time policy or program evaluation using similar causal inference frameworks.

Assessing Campaign Performance Using Chi-Square(A/B) Test for Independence - 🔗 **View project**

- **Context**: A marketing team needed evidence on whether two different mailer designs drove different customer responses.
- **Action:** Designed an experimental framework capturing demographic segments, outreach types, and response rates across multiple regions. Applied Chi-Square Tests for Independence to compare categorical variables (e.g., strategy type vs. response outcome) for statistically significant differences. Summarized findings in **Tableau dashboards** for leadership review.
- **Results:** Determined one outreach strategy significantly outperformed the other in three high-priority demographic **segments**. Findings informed targeted campaign investments and messaging strategies
- **Growth/Next Steps:** Integrated the analysis into a repeatable template for future A/B testing scenarios across programs and regions.

Publication: Exploring Socio-Behavioral Correlates of Metabolic and Inflammatory Risk in Hispanics Living Along the U.S./Mexico Border: A Pilot Study Concomitantly Collecting Survey Data, Blood and Hair Samples, and Physical Measures, by Gabriel A. Frietze, Cai Xu, Bibiana Mancera, Elisa Robles, Escajeda, Alyssa A. Martinez,Michelle Gil, Diana P. Flores, Khodeza Begum, Panfeng Liang, Abhijit Mandal, **Michael Nsiah-Nimo**, Nilotpal Sanyal, Ming-Ying Leung, Michael J. Kenney and Robert A. Kirken