Mauro Florez

+1 (346) 610-0770 | mauroederfr@gmail.com | Bellevue, WA | linkedin.com/in/mauroefr | mauroflorez.github.io

PROFESSIONAL SUMMARY

Data Scientist with a Ph.D. in Statistics and expertise in Bayesian modeling, statistical analysis, machine learning, forecasting, and decision-making under uncertainty. Skilled at translating complex data into actionable insights that drive measurable impact. Collaborative team player with a strong record of delivering high-quality analytical solutions.

EDUCATION

Rice University — **Houston, TX, USA**
PhD, Statistics — May 2025 — GPA: 3.95
MA, Statistics — August 2024 — GPA: 3.94

Universidad Nacional de Colombia — **Bogotá, Colombia**
BS, Statistics — June 2019 — Honors: Ranked #1 in class

Universidad Sergio Arboleda — **Bogotá, Colombia**
BS, Mathematics — September 2017 — Honors: 75% tuition waiver scholarship

SELECTED WORK EXPERIENCE

Amazon — **Seattle, WA, USA**
Data Scientist Intern — June 2025 - Aug. 2025

- Developing an LSTM-based deep learning model in TensorFlow to estimate community sentiment over time, incorporating temporal patterns to improve prediction accuracy.
- Implementing the solution on AWS SageMaker to enable scalable, production-ready deployment and continuous sentiment monitoring.
- Collaborating with applied scientists and engineers to refine model design and ensure alignment with business priorities.

Rice University — **Houston, TX, USA**
Graduate Research Assistant — Aug. 2020 - May 2025

- Developed Bayesian models for multivariate, graphical, and count data, leading to publications in *Bayesian Analysis* and the *Journal of Quantitative Analysis in Sports*.
- Created and published the *MultRegCMP* R package (CRAN) for multivariate count data modeling; additional open-source work shared on GitHub.
- Mentored student teams as a D2K Lab Fellow, guiding machine learning projects including deep learning models for educational performance assessment.

Department of Science, Technology, and Innovation (Minciencias) — **Bogotá, Colombia**
Data Scientist — May 2019 - June 2020

- Managed and optimized institutional databases in SQL to support dashboards, ensuring accurate and timely data publication.
- Created interactive Tableau dashboards for national science and technology metrics, providing government agencies with data-driven insights for policymaking.
- Led end-to-end data consolidation, cleaning, and validation workflows to enhance the quality and accessibility of key performance indicators distributed across Minciencias portals.

SELECTED DATA ANALYSIS PROJECTS

A Bayesian Approach for Inference on Mixed Graphical Models — **Houston, TX, USA**
Rice University — May 2024

- Developed a novel Bayesian graphical model (BMGM) for mixed data types (continuous, categorical, count, zero-inflated) that identifies conditional independencies between variables, providing interpretable network structures even with missing data. Applied to adolescent eating disorders health data, uncovering insights into behavioral changes and treatment outcomes.
- Designed and implemented a custom MCMC algorithm with spike-and-slab priors, achieving superior performance in simulations and real-world health applications.

Forecasting Modeling for Health & Insurance Risk — **Bogotá, Colombia**
Universidad del Rosario — May 2018

- Built a survival analysis model using longitudinal health data to identify worker subgroups at elevated risk of lung disease in one of Latin America's largest open-pit coal mines, leading to targeted preventive health campaigns.
- Developed and deployed a cost forecasting model for insurance claims related to musculoskeletal disorders, enabling the company to project expected costs, flag outliers, and enhance fraud detection across all regions of Colombia.
- Collaborated with interdisciplinary teams to ensure robust model implementation, actionable insights, and stakeholder adoption.

SKILLS

Programming*:* Python (TensorFlow, PyTorch, Scikit-learn), R (package developer), SQL, DataGrip, Tableau, Git
Machine Learning & AI*:* Deep Learning, Large Language Models (LLMs), Autoencoders, Forecasting, NLP, Model Evaluation
Statistics & Modeling: Bayesian Inference, Multivariate Analysis, Graphical Models, Time Series Analysis, Experimental Design
Cloud & Data Management*:* AWS (EC2, S3, SageMaker), Data Engineering, ETL, Visualization (Tableau)
Languages*:* English (fluent), Spanish (fluent), Italian (basic)