

CLÉMENT MUGENZI

New York, NY • cmugenzi01@gmail.com • (xxx) xxx-xxxx • [LinkedIn](#) • [Portfolio](#) • [GitHub](#)

PROFILE

Senior Data Scientist with 5+ years of experience building ML models and causal inference frameworks in healthcare, promoted from Junior to Senior within 2 years for consistent delivery of high-impact causal inference and ML work. Combines deep experimentation expertise (A/B testing, propensity weighting, difference-in-differences) with production AI/ML development on GCP. Seeking roles that apply rigorous experimentation methodology to AI product development.

SKILLS

Programming: Python, R, SAS, SQL, Hive, Hadoop, GCP (BigQuery, Vertex AI)

ML / AI: XGBoost, LightGBM, Random Forest, Logistic Regression, SVM, PCA, CNN, RNN/LSTM (TensorFlow, PyTorch)

Experimentation: A/B Testing, Experiment Design, Power Analysis / MDE Calculation, Sequential Testing, Bayesian A/B Testing, Propensity Score Weighting, Difference-in-Differences, Regression Discontinuity, Instrumental Variables, Two-Way Fixed Effects, Doubly-Robust AIPW, Heterogeneous Treatment Effects (CATE) / Uplift Modeling

Statistics: GLM, Longitudinal Data Analysis, Survival Analysis, Statistical Inference

Libraries: Pandas, NumPy, Scikit-learn, Matplotlib, Tidyverse, TidyModels, R Shiny, R Markdown

AI Tooling: Claude (Anthropic) — Claude Code, hooks & harnesses for model evaluation and testing workflows

WORK EXPERIENCE

Senior Data Scientist

Jun 2023 – Present

Aetna, a CVS Health Company, New York, NY

- Lead a monthly Evaluations Workgroup (since Dec 2024), consulting ~30 data scientists across the org on A/B testing and program evaluation design — advising on study design, causal inference methods, and reporting, improving study design quality across evaluation projects org-wide.
- Led an observational study evaluating Aetna's Integrated Care Management program (n = 25K members) using propensity score weighting and Difference-in-Differences — estimating a statistically significant 3% reduction in total cost of care ($p < 0.05$), informing revised enrollment criteria adopted across the program.
- Built a LightGBM model predicting 90-day inpatient risk for Dual Eligible Special Needs members (AUC 0.70, population of ~60K members); applied SHAP values to reduce Health Risk Assessment dimensionality from 80 to 32 features, surfacing top clinical predictors driving risk.
- Deployed 5 production ML models on GCP using an MLOps framework with Vertex AI Model Registry and scheduled batch scoring pipelines, enabling automated inference across ~500K members per month.
- Leveraged Claude (Anthropic) via Claude Code and evaluation harnesses to accelerate model development workflows — reducing dev cycle time by 50% through automated testing and code review within the data science team.

Data Scientist

Jun 2021 – Jun 2023

Aetna, a CVS Health Company, New York, NY

- Designed and deployed an XGBoost model to identify members most likely to engage with care management calls across a Medicaid population of ~2.3M members, incorporating call history and recent clinical events — increasing program enrollment by 10% and improving nurse outreach efficiency by 30%.
- Developed R Shiny dashboards used by 10 clinical and operational stakeholders to communicate program evaluation results and KPIs, reducing ad hoc reporting requests by 3% and saving approximately 5 hours/week in manual reporting.

Research Assistant

Jun 2020 – Sep 2020

Columbia Heffner Biomedical Imaging Lab, New York, NY

- Extracted, transformed, and analyzed CT imaging data from COPD patients in R; built visualizations including Sankey diagrams to communicate disease progression patterns.
- Conducted A/B tests on contingency tables to assess independence between clinical features, validating results with permutation tests.

- Modeled the association between emphysema subtypes over time using a Multinomial Log-linear model; built a MARS regression model achieving 90% accuracy for percent emphysema prediction.
- Benchmarked MARS against an XGBoost model (82% accuracy) in Python, providing cross-framework comparison.

EDUCATION

M.S., Biostatistics

May 2021

Columbia University, New York, NY

Relevant Coursework: Data Science, Statistical Methods, Statistical Inference, Causal Inference

B.S., Chemistry (Honors)

May 2016

University of Central Arkansas, Conway, AR

PROJECTS

Independent Project: Green Taxi Tip Prediction

- End-to-end ML pipeline in Python: data cleaning, feature engineering, EDA, and model building. Achieved 96.1% AUC with Gradient Boosting for tip classification; followed by Random Forest regression for tip amount (MSE = 0.8).

AWARDS & HONORS

- Taub Institute Award, Columbia University (2019)
- Nicole Wable Hatfield Scholarship (2014)
- Rwanda Presidential Scholar — awarded to top 50 students nationally (2012–2016)