Bernard Frimpong

Data Analyst | Credit Analytics · Statistical Modeling · Forecasting

Akron, OH | +1 (310) 404-8233 | bernardfrimpong614@gmail.com
linkedin.com/in/bernard-frimpong-681008222 | github.com/bernardfrimpong040-pixel

SUMMARY

Data analyst with an M.S. in Applied Mathematics and a portfolio of end-to-end, reproducible projects in credit risk, survival analysis, disease forecasting and machine learning: SQL and Python analysis of **880,000+ commercial loans**, a leukemia survival model scoring **0.754 IPCW C-index** on an unseen hospital, weekly flu forecasts submitted to the **CDC FluSight** challenge, and six rounds of validated gains in a Kaggle competition. Known for honest validation (grouped cross-validation, leakage checks) and for explaining technical results in plain language, from 3+ years of university teaching.

TECHNICAL SKILLS

Programming & Data	Python (pandas, NumPy, SciPy, Matplotlib), SQL (DuckDB), R, MATLAB, Git/GitHub, LaTeX
Machine Learning	scikit-learn, LightGBM, XGBoost, CatBoost; grouped and nested cross-validation, target encoding, model blending, permutation importance, data-leakage checks
Statistics & Modeling	Survival analysis (Kaplan–Meier, log-rank, Cox and penalized Cox, random survival forests; lifelines, scikit-survival), regression, hypothesis testing, time-series forecasting, ODE/compartmental models (SIRS), numerical methods
Credit & Portfolio	Vintage analysis, default and loss rates, concentration limits (HHI), early warning indicators, roll-forwards, peer benchmarking
BI & Reporting	Power BI (DAX), Excel (COUNTIFS, INDEX/MATCH, conditional formatting), PowerPoint

PROJECTS

Commercial Credit Portfolio Analysis | Python, SQL, Power BI, Excel 2026
github.com/bernardfrimpong040-pixel/credit-portfolio-analysis

- Analyzed **882,533** SBA small-business loans ($170B approved) in SQL and Python; found that loans approved in 2005–2008 were 21% of dollars lent but caused **54% of losses**.
- Built an early warning indicator from 2-year default rates that flagged the 2010 loss peak **3 years ahead**, 2 years before the U.S. unemployment rate did (FRED data).
- Designed a formula-driven Excel limit monitor that identified **3 concentration breaches**, and a 4-page Power BI dashboard with DAX measures.
- Benchmarked lenders on an actual-vs-expected basis (same year, loan size and industry), correcting misleading raw rankings; presented findings in a 5-slide executive summary.

Leukemia Overall Survival Prediction | Python, lifelines, scikit-survival, XGBoost Oct 2026
github.com/bernardfrimpong040-pixel/aml-survival-prediction

- QRT × Institut Gustave Roussy challenge (Challenge Data, ENS Paris): predicted overall survival of myeloid leukemia patients from blood counts, karyotype text and gene mutations; scored **0.754 IPCW C-index** on the public leaderboard, a hospital the model never saw.
- Engineered 91 features, including a parser for free-text karyotypes; beat the IPSS-R clinical risk score (0.671) with penalized Cox, random survival forest and XGBoost-Cox models.
- Validated with centre-grouped, nested cross-validation (15 folds); final ensemble reached **0.725** ± **0.021** and beat the best single model on **15 of 15 folds**.
- Derived Kaplan–Meier and the IPCW C-index by hand and matched library output; showed with permutation importance that mutation burden is the strongest predictor; 33-page LaTeX write-up.

CDC FluSight Influenza Forecasting | Python, ODE models, GitHub Oct 2026 – Present **[ONGOING]**
github.com/bernardfrimpong040-pixel/flu-forecasting-sirs

- Submitting weekly probabilistic forecasts of state-level flu hospitalizations to the CDC FluSight 2026–27 challenge; every submission passes the hub's automated validation checks.
- Built a baseline, a damped-trend model and a mechanistic SIRS model (RK4 solver, weekly reproduction number); backtested on 2025–26 with weighted interval score (WIS) by epidemic phase.
- The trend model cut growth-phase error by **17%**; the SIRS model matched the baseline overall with near-zero growth-phase bias, and exposed systematic overshoot at epidemic peaks.

Airline Passenger Satisfaction (Kaggle Playground S6E10) | Python, LightGBM, XGBoost, CatBoost Oct 2026
github.com/bernardfrimpong040-pixel/kaggle-airline-satisfaction

- Raised ROC AUC from 0.95836 to **0.96056** over six documented rounds of feature engineering, fold-safe target encoding and hill-climbing model blends.
- Cross-validation predicted the leaderboard to within 0.0005 every round; a deliberate leakage demo showed how leaky encoding inflates CV by **8**× the real gain.

Traffic State Reconstruction (IEEE Big Data Cup 2026) | Python, Kaggle Oct 2026 – Present **[ONGOING]**
TrafficFlowBench competition on a freeway network; final submission Nov 2026

- Reconstructing missing speed and flow data and predicting queues with a physics-informed approach (LWR conservation law, shockwave speeds); preparing working notes for the IEEE Big Data Cup.

RESEARCH

Do influenza forecasts fail at turning points? | Python, public FluSight hub data In preparation **[ONGOING]**
Phase-stratified evaluation of the CDC FluSight models, 2023–2026 | target: medRxiv preprint, then journal

- Measuring how much worse every FluSight model and the CDC ensemble perform near epidemic peaks than in growth or decline, how late forecasts detect the turn, and how far they overshoot.

Computational Model of Flow Near Permeable Boundaries | Python, MATLAB 2025 – 2026
Master's thesis research, University of Akron

- Built a numerical simulation of Stokes flow using the method of regularized Stokeslets, translating a physical problem into a solvable computational model.
- Validated results against known analytical solutions, analyzed how parameters affected accuracy and convergence, and presented findings in a written thesis and defense.

EXPERIENCE

Mathematics Instructor Aug 2026 – Present
T-Squared Honors Academy-North East Cleveland, OH

- Teach Algebra 1, Algebra 2, and Geometry; track assessment results to identify learning gaps, and built a LaTeX library of curriculum and assessment materials across three courses.

Graduate Teaching Assistant 2024 – 2026
University of Akron, Department of Mathematics Akron, OH

- Led recitations, graded assessments, and held office hours for undergraduate calculus and statistics, explaining quantitative concepts in plain language to students from varied backgrounds.

Teaching Assistant [Year] – [Year]
Kwame Nkrumah University of Science and Technology Kumasi, Ghana

- Assisted with undergraduate mathematics courses, tutorials, and grading.

EDUCATION

M.S., Applied Mathematics 2024 – 2026
University of Akron Akron, OH
Coursework: numerical analysis, mathematical modeling, probability & statistics, differential equations

B.S., Mathematics 2018 – 2022
Kwame Nkrumah University of Science and Technology (KNUST) Kumasi, Ghana