

Donald Jordy Kenfack

DATA SCIENTIST · DATA ANALYST

jordykenfack@outlook.com · +90 555 153 87 11 · Ankara, Turkey · Open to remote

CONNECT ♦ GitHub ♦ LinkedIn ♦ Portfolio

SUMMARY

Mathematics student (B.Sc. completed, M.Sc. in progress) building toward data science and analytics. I take problems from raw data to a defensible decision, framing the question, testing it statistically, modeling it honestly, and communicating the result plainly. Five end-to-end projects, each with tests, documentation, and code public on GitHub, spanning **machine learning**, **time-series forecasting**, **experimentation**, and **analytics engineering**. Comfortable working with AI-assisted development and able to explain every decision.

TECHNICAL SKILLS

Languages	Python, SQL (R, coursework)
Machine Learning	scikit-learn; logistic regression, gradient boosting; baseline-first comparison; cross-validation & temporal validation; leakage prevention; class imbalance; precision@k, ROC-AUC, log-loss, Brier; calibration and error analysis; feature engineering
Statistics	Hypothesis testing (chi-square, Welch t-test, two-proportion z-test); A/B evaluation; effect sizes (Cramér's V, Cohen's d); regression; exploratory data analysis
Time Series	Walk-forward backtesting; leakage-safe lag/rolling features; forecast metrics (MAE, WAPE)
Data & Viz	pandas, NumPy, Matplotlib, Plotly, Jupyter, Streamlit; cleaning messy real data; star-schema modeling; KPI design
Workflow	Git & GitHub, pytest, reproducible pipelines, virtual environments; SQLite; LLM-assisted development

PROJECTS

- Customer Churn Prediction: telecom retention

github.com/jordykenfack/telco-churn

 - Framed churn as a **ranking problem under a real constraint** (~200 retention calls) on the public IBM Telco dataset (7,043 customers); evaluated on **precision-in-top-200**, not accuracy.
 - Reached **74% precision in the top 200** vs 63.5% for the best heuristic and 24.5% random (ROC-AUC **0.846**); the simpler logistic model beat gradient boosting, with leakage-safe pipelines, calibration, and error analysis. Python, scikit-learn, SciPy, pytest.
- NBA Game Prediction: can features beat Elo?

github.com/jordykenfack/nba-game-prediction

 - Predicted winners over **63,000 games (1947–2015)** with a **temporal split** and leakage-safe rest/form features, benchmarked against FiveThirtyEight's Elo.
 - Model **matched Elo at ~67% accuracy** but honestly did not beat it, baseline rigor and probabilistic evaluation (log-loss, Brier, calibration). Python, scikit-learn, pandas.
- Demand Forecasting: 7-day-ahead bike-share rentals

github.com/jordykenfack/demand-forecasting

 - 7-day-ahead forecast with **walk-forward backtesting** (517 out-of-sample forecasts) and strictly leakage-safe features on real UCI data; best model beat the moving-average baseline by **27% (MAE)**, and I quantified a ~17% gain from a weather-forecast feed. Python, scikit-learn.
- Revenue & Retention Analytics: subscription business (SQL + experimentation)

github.com/jordykenfack/nordwind-analytics

 - Advanced SQL (window functions, CTEs, cohorts) on a 5-table warehouse (synthetic, disclosed); found paid-channel customers worth **~40% less** and evaluated a win-back campaign with a **two-proportion z-test** (+2.1pp, p≈0.002). Delivered a one-page executive summary.
- Retail Operations Dashboard: UK online retailer (BI)

github.com/jordykenfack/retail-dashboard

 - Cleaned **1,067,371 real invoice lines** through an audited pipeline; shipped a unit-tested KPI layer and a 4-view Streamlit/Plotly dashboard with RFM segmentation (5,852 customers). pandas, Streamlit.

EDUCATION

M.Sc. Mathematics, Middle East Technical University (ODTÜ), Ankara · 2025 – present
B.Sc. Mathematics, secondary field in Data Analytics, TED University, Ankara · 2021 – 2025
Relevant coursework: probability & statistics, regression, linear algebra, data analytics.
Languages: English · French · Turkish