

Ibrahima Barry

Data Scientist | iblouse49@gmail.com | 614-312-4440 | [LinkedIn](#) | [GitHub](#) | [Portfolio](#) | Brooklyn, New York

Summary

Data Scientist with more than 5 years delivering analytics and machine learning solutions from problem definition through deployment. Built forecasting, anomaly detection, and predictive models that improved conversion and retention and generated \$400K+ in quantified annual savings. Experienced translating business questions into measurable analyses and communicating results through dashboards and written recommendations.

Skills

- **Programming:** Python, SQL (Redshift, BigQuery), PySpark, R, PyTorch, TensorFlow
 - **Machine Learning & AI:** XGBoost, Random Forest, Neural Networks, Ensemble Methods, K-means Clustering, Time Series Forecasting, LLM Integration, Retrieval Augmented Generation
 - **Cloud & Tools:** AWS (S3, Glue, Redshift, SQS), Google Cloud (BigQuery, Vertex AI), Databricks, Snowflake, MLflow, Docker
 - **Analytics & Statics:** EDA, KPI reporting, A/B testing, ANOVA, chi-square tests, Multivariate Analysis
 - **Visualization & Reporting:** Tableau, Power BI, SAS Visual Analytics, executive dashboards, KPI reporting

Experience

Senior Data Science Consultant *Data In Motion* **New York, NY Nov 2024 – Present**

- Led analytics delivery from problem definition to deployment, improving conversion rates by 18% using funnel and survey analysis.
 - Built and deployed time series forecasting models in Python to optimize webinar scheduling, improving resource allocation efficiency by 20%.
 - Delivered competitor analysis across 8 platforms and translated findings into retention actions that improved customer retention by 15%.

Analytics Engineer **USA for UNHCR** **New York, NY** **May 2022 – Jan 2024**

- Implemented an ML-driven anomaly detection and alerting workflow for donation approvals using AWS SQS, preventing fraudulent transactions and saving \$400K annually.
 - Built a production Redshift pipeline using AWS Glue to integrate Salesforce and S3 data sources, reducing manual processing by 35% through automated ETL and complex SQL.
 - Developed gradient boosting models with engineered features, improving campaign donations by 20% lift.
 - Applied k-means clustering for customer segmentation, creating targeted retention strategies visualized in Tableau dashboards to guide cross-departmental digital campaigns.
 - Collaborated with agency management to report on application efficiency via Tableau dashboards, directly informing executive decision-making.

Data Science Fellow Sharpest Minds New York, NY Mar 2021 – Mar 2022

- Developed predictive models to assess Type 2 diabetes risk using CDC BRFSS data, aiding early diagnosis.
 - Processed and cleaned datasets with over 400,000 rows and 275 columns, prioritizing sensitivity to achieve an 83.77% sensitivity rate with logistic regression.
 - Identified 25 key predictors of Type 2 diabetes, including BMI, general health, and income level.

Data Analyst *Self-Employed* *New York, NY Feb 2013 – Feb 2021*

- Forecasted seasonal product demand using **ARIMA**; improved winter sales by **15%**, reduced unsold stock by **10%**.
 - **Increased weekly avocado orders by 25%** and adjusted deliveries to Thursday, resulting in a **12% sales boost** and reduced stockouts, after identifying a 30% demand spike on Fridays and Saturdays.

Projects

Proactive Churn Prevention System | Python, AI Agents, Lifelines and Vertex AI | [Repo](#) | [Medium](#) | [Portfolio](#)

- Built dual-model architecture (classification + survival analysis) to predict churn timing, identifying \$2.54M in at-risk CLV and deriving a data-driven intervention window (Day 45-95)
 - Validated interventions through 5-variant A/B testing with Bonferroni correction, achieving 54.4% churn reduction ($p < 0.0001$) and 158.8x ROI across channels

- Developed multi-agent system using Google ADK to operationalize retention workflow with automated behavioral monitoring, risk scoring, and channel selection

Multi-Agent Loan Processing System with RAG Underwriting | Python, Gemini, Transformers | [Repo](#) | [Medium](#) | [Kaggle](#)

- Designed multi-agent Gen AI system for loan processing workflow including KYC verification, financial data extraction, risk scoring, and transparent underwriting via Retrieval Augmented Generation
- Generated synthetic financial data for privacy-compliant demonstrations and implemented comprehensive financial calculations for realistic loan assessments
- Demonstrated end-to-end deployment of LLM-based business application

Fraud Detection Model and MLOps Deployment | Python, MLFlow, Streamlit | [Repo](#)

- Built and deployed end-to-end ML model for fraudulent transaction detection from EDA through production deployment
- Integrated MLflow for experiment tracking and Streamlit for real-time fraud detection interface
- Created user-friendly application demonstrating full ML lifecycle management

Education

Master of Statistics & Applied Mathematics

CUNY Hunter College, New York, NY

Certifications

• AWS Certified Machine Learning – Specialty	Mar 2024
• AWS Certified Cloud Practitioner	Feb 2024
• SAS Certified Clinical Trials Programmer Using SAS 9	Dec 2020
• SAS Certified Specialist: Forecasting and Optimization Using SAS Viya 3.5	Sep 2020
• SAS Certified Specialist: Machine Learning Using SAS Viya 3.4	Aug 2020
• SAS Certified Base Programmer for SAS 9	Mar 2018

Languages

French

Fluent