

Massoudou Namata Abdoul Wahid

Data Scientist | Quantitative Analyst

Montréal, QC, Canada | +1 438 994 6004 | massaoudouabdoulwahid@gmail.com

[LinkedIn](#) | [GitHub](#) | [Portfolio](#) | French (Native) | English (Fluent)

PROFILE

Data scientist and quantitative analyst (World Bank, CIRANO, Orange Innovation). Designs and maintains **structured, reproducible pipelines** (Python, Git) to clean, reconcile, and consolidate heterogeneous and often incomplete administrative microdata, using **fuzzy matching, record linkage**, and data-quality control. Complementary skills in NLP, time-series forecasting (Bayesian VAR, conditional forecasts), and causal inference.

TECHNICAL SKILLS

Data Engineering: Record linkage, fuzzy matching, deduplication, missing-data imputation, data-quality control, reproducible & version-controlled ETL pipelines (Git), PDF/web scraping (`pdfplumber`), OCR, REST APIs.

Languages & Tools: Python (`pandas`, `NumPy`, `scikit-learn`, `statsmodels`), SQL, R, Git, Power BI, $\text{\LaTeX}$.

Quant. & ML Methods: Semantic similarity, transformer models (BERT/FinBERT), generative LLMs, XGBoost, random forests, LASSO/Ridge/Elastic Net, time series (Bayesian VAR), DiD/IV/RDD.

PROFESSIONAL EXPERIENCE

Research Consultant (Data Science), Development Impact (DIME), World Bank Group

Mar. 2026 – Present

- Developed automated PDF extraction pipelines (`pdfplumber`) converting scanned documents into structured datasets.
- Designed and maintained reproducible Python pipelines processing administrative microdata (business registry, tax records) for two countries.
- Built fuzzy matching and record linkage procedures to deduplicate and consolidate firm records across heterogeneous source systems with no common unique identifier.
- Reconstructed workers' tax identifiers (TIN) from fragmented, incomplete sources via exact and fuzzy matching (name, phone) and missing-value imputation.
- Field mission to establish data partnerships with public and private institutions: counterpart identification, collaboration framework, and signing of data-sharing agreements.
- Wrote the final deep-dive diagnostic report on the Liberia and The Gambia labour markets, drawing on multiple primary and secondary data sources.

Data Analyst, Orange Innovation, Paris, France

Mar. – Aug. 2024

- Built an OCR pipeline automating PDF data extraction, cutting manual processing time by over 80%.
- Processed and cleaned large-scale administrative databases (millions of records) in Python and R, on incomplete and inconsistent data.
- Developed a concept note on impact analysis of Orange's offers: literature review and design of a full impact-evaluation methodology based on DiD and RDD techniques.

Research Assistant, Chair in Macroeconomics and Forecasting, ESG UQAM

Jan. 2025 – May 2026

- Real-time maintenance and updating of the macroeconomic time-series database (Bank of Canada: policy rate, inflation expectations, output gap); analysis in R and MATLAB.
- Contributed to a *working paper* on media sentiment and monetary policy: estimated the Bank of Canada Taylor rule (GMM/IV, sentiment treated as endogenous) and identified a Bayesian SVAR via sign restrictions, with sentiment indicators built from 50,000+ news articles (transformer models FinBERT, generative LLMs); econometric tables and impulse-response functions.

Quantitative Research Assistant, CIRANO, Montréal

Sept. 2025 – May 2026

- Scraped public budget plans and built a database of revenue and expenditure items with realized and forecast values by year (T, T+1).
- Computed risk measures (Value-at-Risk) on realized-versus-forecast errors to calibrate the government's budgetary risk provisions.
- Wrote a *policy report* on setting provisions and formalized a provisioning model for a public decision-maker.

EDUCATION

M.Sc. Economics (quantitative), ESG UQAM, Montréal, Canada

2024 – 2026

M.Sc. Development Economics, CERDI, U. Clermont Auvergne, France

2022 – 2024

B.Sc. Economics & Management, U. Hassan II, Casablanca, Morocco

2018 – 2021