

# Roy Ma

## Citizen Data Scientist

[[roy.ma9@gmail.com](mailto:roy.ma9@gmail.com)] · [[linkedin](#)] · Manassas, VA 20110

Citizen Data Scientist with 4 years of applied data engineering and 10 years in finance. Hands-on with production-grade ETL and machine learning for legal, tax, and financial domains. Owned pipelines powering capital reconciliations, jurisdictional forecasting, and entity workflows. Built audit-ready systems leveraging GCP, AWS, Airflow/Composer, Spark, and dbt. Designed statistical experiments and deployed ML/DL solutions using Scikit-learn, PyTorch, and TensorFlow. Experienced with LLM workflows (RAG, fine-tuning) using LangChain, Transformers, and SageMaker. Skilled in data storytelling and stakeholder communication via Tableau. Core technical toolkit includes Python, SQL, Git, and Docker. Coached internal teams on adopting data-driven workflows to support critical business decisions. Does not require U.S. visa sponsorship.

## Professional Experience

---

### Citizen Data Scientist, SSR (守税人)

2020 to 2025, Beijing, China

- Owned ETL lifecycles driving regulatory reporting, entity workflows, and capital reconciliations; automated high-risk, time-sensitive deliverables using `GCP`, `BigQuery`, `Composer/Airflow`, and `Spark/PySpark`, eliminating manual errors and enabling reusable pipelines.
- Developed and deployed machine learning and deep learning logic using `Scikit-learn`, `PyTorch/TensorFlow` to classify transactions, estimate positions, and forecast jurisdictional exposure; reduced client validation cycles by 10+ hours per engagement.
- Translated domain expertise from legal, finance, and tax teams into scalable data logic; deployed internal tooling and dashboards via `Snowflake`, `Tableau`, and `Docker`, reducing rework tickets by 40%.
- Designed validation, monitoring, and audit layers for production pipelines to ensure reliability, traceability, and reproducibility across systems.
- Coached non-technical stakeholders on internal tooling, integrating data-driven workflows into advisory pipelines and client-facing deliverables.

### Director of Business Development, Cleveland International Fund

2017 to 2019, Cleveland, OH, U.S.

- Directed investor relations for a portfolio of 300+ institutional and private investors, overseeing investor reporting, capital calls, and regulatory compliance with meticulous attention to data accuracy and transparency.
- Raised \$250M+ in direct capital commitments, personally originating and closing repeating deals with 35+ key LPs, and leading capital formation strategies through fund marketing and roadshows, sourced and secured commitments from 60+ new investors.
- Synthesized complex market data and financial models into clear, actionable narratives, delivering high-impact storytelling to investors and senior leadership.

### Head of International Department, MultiBank Group

2016 to 2017, Beijing, China

- Oversaw a 26-member international department, directing business development, account management, and international operations strategy for a 350+ employee global brokerage.
- Generated \$8B+ in daily trading volume, managing 120,000+ active global retail clients and personally overseeing 50+ key wealth management clients, with \$400M+ in assets under management.

- Directed global marketing, integrating predictive analytics via machine learning to drive targeted client acquisitions and optimize incentive programs, resulted in \$120M+ in YoY incremental revenue.
- Improved Salesforce CRM, personally contributing custom codes that enabled smart client segmentations, precise lead scoring, and targeted marketing campaigns.
- Owned departmental KPI setting, performance evaluations, and reporting directly to the CEO, with full authority over recruitment, team expansion, and operational oversight.

## Projects

---

### Kaggle Competitions Expert, top 0.8%

LLMs - You Can't Please Them All (Silver Medal, Ranked Top 3%)

- Designed and implemented LLM prompt optimization pipelines using `LangChain` for workflow orchestration, `Transformers` for model interfacing, and `scikit-learn` for performance analysis.
- Addressed critical business challenges in LLM reliability and prompt tuning workflows automation, reducing manual prompt engineering cycles, improving accuracy-to-cost ratio in enterprise-scale LLM deployments.

Santa 2024 - The Perplexity Permutation Puzzle (Bronze Medal, Top 5%)

- Developed and applied optimization algorithms, including `evolutionary` strategies and `gradient descent` methods, to solve complex black-box search problems in natural language processing tasks.
- Demonstrated a structured approach to solving open-ended problems, translating complex algorithmic results into actionable insights and scalable solutions for real-world business decision-making.

## Education

---

Master of Liberal Arts (ALM), Data Science, Harvard ES (In Progress) MicroMasters,  
Statistics and Data Science, MIT Professional School  
Bachelor of Arts, Economics, University of Maryland Baltimore County

## Certificates and Languages

---

English - Native, Mandarin - Native, Japanese - Business Conversational  
FINRA Series 63 - Uniform Securities State Law, FINRA Series 82 - Private Securities Offerings Representative

[\[Codes\]](#) · [\[Kaggle\]](#) · [\[Github\]](#)