Roy Ma

Citizen Data Scientist

[**roy.ma9@gmail.com**] · [**linkedin**] · Manassas, VA 20110

Citizen Data Scientist with 5 years in applied data engineering and 10 years in finance. Experienced in ETL in production and machine learning across legal, tax, and financial domains. Owned pipelines powering capital reconciliations, jurisdictional forecasting, and entity-level workflows. Built data platforms on GCP and AWS with Airflow/Composer, Spark, and dbt. Designed statistical experiments and deployed ML/DL models using Scikit-learn, PyTorch, and TensorFlow. Hands-on with LLM customization workflows (RAG, fine-tuning) via LangChain, Transformers, and SageMaker. Strong in data storytelling and stakeholder communication using Tableau. Coached teams to integrate data-driven practices into business decision cycles. Core stack: Python, SQL, Git, Docker. U.S. national, no sponsorship needed.

Work 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.
- Built and deployed machine learning and deep learning logic using `Scikit-learn`, `PyTorch/TensorFlow` to classify transactions, estimate positions, and forecast jurisdictional exposure; reduced validation cycles by 10+ hours per engagement.
- Converted domain expertise from legal, finance, and tax teams into scalable data logic; deployed internal tooling and dashboards via `Snowflake`, `Tableau`, and `Docker`, cutting 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 deliverables.

Director of Business Development, Cleveland International Fund
2017 to 2019, Cleveland, OH, U.S.

- Directed investor relations for 300+ institutional and private investors. Oversaw investor reporting, capital calls, and regulatory compliance with meticulous attention on accuracy and transparency.
- Raised $250M+ in direct capital rounds; originated and closed repeat deals with 35+ key LPs, and led fund marketing and roadshows to secure 60+ new investors.
- Translated complex market data and financial models into clear, actionable narratives. Delivered insightful storytelling to investors and senior leadership, highlighting critical messages.

Head of International Department, MultiBank Group
2016 to 2017, Beijing, China

- Led a 26-member international team across business development, account management, and operations for a 350+ employee global brokerage.
- Managed $8B+ in daily trading volume across 120,000+ global retail clients; personally oversaw 50+ high-value clients with $400M+ in AUM.
- Directed global marketing strategy and client acquisition; applied machine learning to

optimize incentive programs, driving $120M+ YoY incremental revenue.

- Enhanced Salesforce CRM with custom codes to enable smart segmentation, lead scoring, and targeted outreach cycles.
- Set departmental KPIs, led performance reviews, and reported directly to the CEO; owned team expansion, hiring, and overall operational execution.

Selected Projects

Kaggle Competitions Expert, top 0.8%

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

- Built LLM workflows using `LangChain` (workflow orchestration), `Transformers` (model interface), and `scikit-learn` (iterative metrics).
- Tackled core challenges in SOTA LLM automation pipelines. Cut manual engineering cycles and improved accuracy-to-cost efficiency in production deployments.

`Santa 2024 – The Perplexity Permutation Puzzle` (Bronze Medal, Ranked Top 5%)

- Applied optimization techniques including `evolutionary` strategies and `gradient descent families` to solve black-box search problems in NLP.
- Translated complex algorithmic solutions into clear, scalable, and actionable insights to support real-world business decision-making.

Education

Master, Data Science, Harvard Extension School (Expected 2027)
MicroMasters, Statistics and Data Science, MIT Professional School
Bachelor, Economics, University of Maryland, Baltimore County

Languages & Certifications

English - Native, **Mandarin** - Native, **Japanese** - Business Conversational
FINRA Series **63** - Uniform Securities State Law
FINRA Series **82** - Private Securities Offerings Representative
U.S. National, Does **not** require U.S. visa sponsorship

Technical Skills

Core: Python, SQL, Git
Platform: GCP, AWS
Data Engineering: Databricks, Spark, dbt, Airflow, BigQuery, Docker
Frameworks: PyTorch, TensorFlow, scikit-learn, LangChain, Transformers
Visualization: Tableau

[**Codes**] · [**Kaggle**] · [**Github**]