

THALYANN OLIVO

Data Scientist • AI Engineer

New York, NY • thalyann.olivo@gmail.com

thalyanns.vercel.app • linkedin.com/in/thalyann-olivo • github.com/tali0n-git

SUMMARY

Data scientist who builds and validates the pipelines behind reporting people actually trust. Currently a Data Science Apprentice at 718 Digital Labs, working in Agile alongside engineers and product partners on Python pipelines and applied AI features. Track record of owning data quality end to end (schema standards, anomaly detection, documentation) and translating technical work into terms non-technical stakeholders can act on. Building AWS and PySpark depth for large-scale data work.

TECHNICAL SKILLS

Data & Pipelines: Python (Pandas, NumPy, Scikit-Learn), SQL, multi-source pipeline development, data ingestion & transformation, schema design, KMeans clustering

Data Quality & Governance: Data validation, anomaly detection, audit trails, role-based access control, data dictionaries & documentation

Cloud & AI: AWS Bedrock (hands-on), AWS S3 / Glue / Redshift (in study, AWS Skill Builder), RAG pipelines, prompt design, embeddings, vector DBs

Ways of Working: Agile / Scrum, Jira, code review, FastAPI, SQLAlchemy, React + Vite, REST APIs, Git/GitHub

In progress: PySpark for large-scale processing — public build log at thalyanns.vercel.app/notebook

Languages: English, Russian, Spanish

EXPERIENCE

718 Digital Labs • Data Science Apprentice • Bronx, NY July 2026 – Present

- Build data pipelines and write analyses in Confluence/Jira to answer live product questions for internal platforms
- Contribute to LLM integrations (AWS Bedrock), prompt engineering experiments, and applied AI agent work on production projects
- Work in Agile with engineers and product partners (standups, sprint planning, and code review); feedback from this process drove the schema discipline and evaluation-logging standards I now apply across my own projects

Prometheus • Data Scientist • Remote Sept 2025 – Dec 2025

- Led a team of 4 delivering standardized outcome measures for an education nonprofit, bridging the client's program staff and the technical team to turn requirements into a working data solution
- Built multi-source Python and SQL pipelines processing intake demographics, attendance, and outcome data; enforced consistent schema and extraction standards so downstream reporting stayed reliable
- Identified and resolved data quality issues across sources, then delivered weekly reporting and a final documented handoff signed off by the client's Executive Director

Stony Brook University • Office Assistant • Stony Brook, NY May 2021 – Dec 2022

- Migrated 6 years of legacy records into a standardized digital system, cutting retrieval time by 70%
- Wrote process documentation and trained 3 staff on internal systems, reducing onboarding from 4 weeks to 2

PROJECTS

NYC School Performance Analysis (GitHub: [NYC_Public_Schools_Analysis](#)) Dec 2025 – Present

- Multi-source Python pipeline joining NYC OpenData school metrics with MTA transit data; KMeans clustering to surface how transit access tracks with student outcomes

ALVI-AI — RAG-Powered AI Assistant (alvi-ai.com) April 2026 – July 2026

- Built end-to-end: custom scraper feeding a vector knowledge base, full RAG pipeline from semantic chunking through context injection, and a system-prompt architecture governing scope and safety
- Instrumented evaluation logging to catch retrieval failures before feature work

Capsule — Healthcare Data Platform (GitHub: [capsule_prototype](#)) 2025 – Present

- Full-stack platform with role-based access control, server-side audit trails, built for an LLM summarization layer — FastAPI, SQLAlchemy, React/Vite

EDUCATION

The Knowledge House — Innovation Fellowship, Data Science Track December 2025

- 9-month intensive in Data Science, ML, and Analytics; led backend data infrastructure at the 2025 Bloomberg Hackathon

Kingsborough Community College — A.S., Computer Science • GPA: 3.8 December 2016

NYU Polytechnic Institute — B.S. Electrical & Computer Engineering (in progress) 2013 – 2015