

James Beavan

| beavan@gatech.edu | [linkedin.com/in/jamesbeavan](https://www.linkedin.com/in/jamesbeavan) | github.com/beavan-james

SUMMARY

Junior at Georgia Tech (Mathematics, Data Science concentration) specializing in Data Engineering and Infrastructure, with applied machine learning experience building predictive/ranking models. Skilled in architecting data pipelines, designing database schemas, and building backend APIs using Python, SQL, Dagster, and cloud infrastructure.

TECHNICAL SKILLS

Languages: Python, SQL, Java, Go

Data Infrastructure & Cloud: AWS (S3, EC2, Route 53), Oracle Cloud (OCI), Databricks, PostgreSQL, MySQL, T-SQL, SQLite

Pipeline & Orchestration: Dagster, DuckDB, Docker, Linux/Bash, Git, RESTful APIs

Frameworks & Analytics: Pandas, NumPy, Sklearn, Dash, Reflex, Tableau, Jupyter, Excel

WORK EXPERIENCE

Data Insights Intern

Jun. 2026 - Aug. 2026

Northrop Grumman

Falls Church, VA

- Developed Python scripts to automate documentation generation for the manufacturing analytics pipeline
- Diagnosed and resolved conflicting aggregation logic across two dashboard views, aligning metric definitions and eliminating reporting discrepancies
- Optimized core Tableau dashboard, increasing performance by 25%
- Won CIDO Intern Hackathon developing an app to track and filter international procurement data for orders of over 3 million dollars
- Built a two-stage documentation chatbot using langchain-openai with dynamic file routing and context filtering, optimizing prompts and token allocation to eliminate context pollution

Software & Data Intern

May 2025 - May 2026

Lindaben Foundation

Columbia, MD

- Analyzed Excel-based time logs using Python, Pandas, and Openpyxl to detect 300+ data entry errors and generate visual heat maps, improving entry accuracy by 90%
- Built and documented 30+ REST APIs for users, schools, vendors, deliveries, and orders, including filtering, pagination, and sorting via query parameters
- Implemented role-based access control defining permissions for admins, accounting, vendors, schools, and operations teams
- Developed a GPT-4-assisted migration tool to convert legacy Excel files with 200+ delivery rows to JSON formatted data involving numerous database relationships
- Built and deployed an automated data pipeline using Dagster, DuckDB, and Docker on AWS EC2. Pipeline ingests CSVs via a Dash frontend, performs incremental loads with timestamp-based deduplication, and produces seven aggregation tables utilized in an analytics dashboard. Processes 180K records.

Data Science Intern

Sep. 2024 - Jan. 2025

Intechgambit

Falls Church, VA

- Tested and optimized a Python/Biopython-based tool for retrieving gene sequences from NCBI, identifying its upper limit at 40,000 entries before failure
- Evaluated AI models (ChatGPT, Gemini, Copilot) by testing 50+ prompts; recorded accuracy and runtime metrics to compare performance across tasks

PROJECTS

Stockidence | <https://github.com/beavan-james/Stockidence>

Aug. 2026 - Present

- Built an on-demand stock-confidence platform on Dagster and DuckDB, incrementally ingesting prices, fundamentals, and news sentiment from 4 external APIs, gated by per-endpoint staleness TTLs
- Designed a DuckDB medallion architecture, with staging tables handling data cleaning and mart tables aggregating technical indicators that feed a stock ranking model
- Trained an XGBoost LambdaMART (rank:ndcg) learning-to-rank model over quarterly cohorts with walk-forward validation (rank IC +0.163; top-quintile excess +2.99 pp/qtr)
- Deployed on Oracle Cloud Ampere A1 compute with cron-pull deploys

EDUCATION

Georgia Institute of Technology | Atlanta, GA

Jul. 2025 - May 2028

B.S. in Mathematics, Data Science Concentration

GPA: 3.6/4.0

- Relevant Coursework: Data Structures & Algorithms, Object-Oriented Programming, Mathematics of Data Science, Statistics and Probability, Discrete Mathematics