

Autumn G. Smith

[LinkedIn](#) · [GitHub](#)

Physics-trained engineer building production data systems end to end since 2022. Handed open-ended problems and asked to determine what's feasible, then build it. Currently developing depth in computer vision, with the goal of applying machine learning to remote sensing and scientific data.

Education

Wellesley College, Wellesley, MA — *B.A. in Data Science with Physics Concentration*

May 2022

- Coursework through intermediate physics including Electricity & Magnetism, Quantum Mechanics, and Experimental Techniques
 - Quantitative and computing coursework: Machine Learning, Causal Inference, Linear Algebra, Data Structures, and Databases with Web Interfaces
 - Sigma Xi National Scientific Research Honor Society
 - National Hispanic Recognized Scholar
-

Certifications

dbt Certified Developer — dbt Labs · Issued December 2025 · Valid through December 2027

Research

Research Assistant — Computer Science Department, Wellesley College

PI: Dr. Christine Bassem · Apr 2021 – May 2022

Constrained clustering for balanced group assignment under multiple real-world constraints.

- Implemented a **constrained k-means clustering algorithm from scratch in Python and NumPy** to partition individuals into balanced groups under simultaneous constraints (geographic proximity, schedule compatibility, dietary requirements) — a problem standard k-means cannot solve, since it optimizes cluster compactness with no cardinality constraint.
- Designed an **iterative reassignment heuristic**: points farthest from an over-capacity cluster's center are reassigned to their next-best feasible cluster, cascading through displaced points until all size constraints are satisfied.

- Engineered a **mixed-type feature space**, one-hot encoding categorical attributes alongside continuous geographic features to produce a single usable distance metric.
- Empirically evaluated candidate approaches to select the best-performing method; presented results as a **poster at the SUNY Undergraduate Research Conference**.
- Generalized the tool beyond its original use case to constrained group formation broadly, including instructor-led study group assignment.

Research Assistant — EECS Department, Massachusetts Institute of Technology

PI: Dr. Amar Gupta · Jun 2020 – Aug 2020

- Investigated **FDA policy governing the use of artificial intelligence in telemedicine** and clinical settings, assessing how existing regulatory frameworks accommodate statistical models deployed in high-stakes environments.
- Conducted a literature review of HIPAA's effectiveness, identifying where the framework holds and where it is vulnerable to modern computational and inferential capability.

Introductory Statistics Tutor — Mathematics Department, Wellesley College

Sep 2021 – May 2022

- Tutored a class of 25 students weekly, scaffolding instruction to individual needs and coordinating with faculty on course content.

Professional Experience

Senior Analytics Engineer — All Star Healthcare Solutions

Feb 2026 – Present

- Serve as the organization's **technical scout**: handed open-ended problems and early-stage ideas and asked to determine what is feasible by scoping the question, prototyping approaches, and recommending a path forward to leadership.
- Designed, built, and shipped a complete modern data platform in four months (Fivetran ingestion, BigQuery warehouse, Omni serving layer), including all data modeling, testing, and monitoring; on call for the system.
- Architected a cross-team engineering analytics platform unifying ticket data across heterogeneous team workflows, building a semantic layer that normalizes disparate schemas into comparable metrics for velocity and cycle time.
- Upskilled colleagues in dbt best practices and the conventions behind our data platform.

Analytics Engineering Specialist — HDI Global Insurance Company

Mar 2025 – Feb 2026

- Built a **statistical sampling framework to validate model accuracy**, running repeated sample tests to demonstrate performance consistently above the organization's risk-tolerance threshold and to verify pipeline and modeling refactors.
- Co-led architecture and implementation of the foundational data solution for Guidewire PolicyCenter, designing data models and integrations aligned to national and global reporting requirements.
- Co-led the refactor of a legacy Azure environment into a modern stack on dbt and Snowflake.
- Established repository structure, environment configuration, testing standards, and CI/CD practices across the analytics engineering team.
- Partnered with global engineering teams, including the Home Office in Germany, on API-driven master data integrations harmonizing U.S. datasets to ISO-aligned standards.

Analytics Engineer — Paradox AI

Jun 2022 – Mar 2025 · Promoted from Data Analyst

- Built proof-of-concept implementations in BigQuery and Snowflake evaluating candidate reporting infrastructure; secured C-suite approval to adopt.
 - Optimized warehouse performance in Redshift through partitioning and clustering strategy, substantially reducing transformation runtimes.
 - Refactored legacy models in dbt, separating source from business transformations into a star schema and semantic layer.
 - Diagnosed root causes of complex data quality failures using a forensic, hypothesis-driven approach.
-

Projects

EuroSAT Satellite Image Classifier

github.com/autumngsmith/eurosat

- Fine-tuned a **ResNet18 CNN on the EuroSAT Sentinel-2 dataset** (27,000 images, 10 land-use classes) for land cover classification, reaching **97.63% test accuracy** which is within the range of the published benchmark for the dataset.
- Implemented the training pipeline independently for a freeze-then-fine-tune transfer learning model: 70/15/15 split, resizing and random flip augmentation, and normalization to ImageNet channel statistics so pretrained weights stayed usable.

Personality Classification — Kaggle Practice Competition

github.com/autumngsmith/kaggle_personality_classifier

- Built a **regression-based missing-data imputation strategy**: for each of five continuous fields, fit an imputation model with a transformation chosen from the shape of that field's distribution, then benchmarked it against mean-fill and median-fill on held-out validation data.
- Scored **0.9684**, against a majority-class baseline of roughly 74% Extrovert.

Technical Skills

Languages — Python (NumPy, pandas, scikit-learn), SQL (BigQuery, Snowflake, Redshift, SQL Server, MySQL), dbt, R

ML & Analysis — Clustering and unsupervised methods, statistical sampling and validation, feature engineering

Cloud & Infrastructure — Google Cloud Platform (BigQuery, GCS, Cloud Functions, IAM), Snowflake, Azure, AWS Redshift, Fivetran

Tools — Git/GitHub, Jira, Confluence, Azure DevOps, dbt Cloud, Omni, Qlik, Looker

Spoken — English (native), Spanish (intermediate, B1)