Autumn G. Smith

LinkedIn · GitHub · hello@autumngsmith.com

Physics-trained analytics engineer building production data systems end to end since 2022: governance, ingestion, warehouse, modeling, and testing. Typically 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

Professional Experience

Senior Analytics Engineer — All Star Healthcare Solutions

PE-backed healthcare staffing platform · Feb 2026 – Present

- Serve as the organization's **technical scout**. I am 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.
- As lead analytics engineer, **designed, built, and shipped a complete modern data platform in four months** (Fivetran ingestion from Salesforce and Sage, BigQuery warehouse, dbt transformation, Omni serving layer), including all data modeling, testing, and monitoring, enabling conversational analytics on the semantic layer for business teams including Sales, Finance, and HR; on call for the system.
- **Own the data model end to end**, designing it to scale: built several config-driven dbt models (Jinja macros + seeds) so extending them is a configuration change rather than new code, e.g., a daily fact table combining 14 metrics from 11 source models for the sales funnel, where adding a metric is a one-row change.
- Upskilled colleagues in dbt best practices and the conventions behind our data platform.

Analytics Engineering Specialist — HDI Global Insurance Company

Mar 2025 – Feb 2026

- Validated an undocumented monthly report for external audit, which it passed. Built a repeatable sampling framework to set the grain, clean the data, and surface edge cases, then reused it to verify pipeline and model refactors.
- Led scoping and dbt design for key components of a refactor from a legacy Azure environment to dbt and Snowflake; upskilled the team in dbt best practices.
- Mapped U.S. source data to ISO-aligned master data standards for API-driven integrations with the Home Office in Germany.

Analytics Engineer — Paradox AI

Jun 2022 – Mar 2025 · Promoted from Data Analyst

High-growth conversational AI SaaS unicorn (acquired by Workday, 2025); supported clients including McDonald's, Unilever, and Sodexo.

- Evaluated replacements for the legacy warehouse by building proof-of-concept implementations in BigQuery and Snowflake; recommended BigQuery, which the C-suite approved.
- Migrated transformation workloads into BigQuery, designing a partitioning and clustering strategy that cut runtimes from 45–60 minutes to under 15.

- 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 accuracy on the Kaggle leaderboard**, against a majority-class baseline of roughly 74% Extrovert.

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.
- 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**.

Research Assistant — EECS Department, Massachusetts Institute of Technology

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

- Researched **FDA policy on AI in telemedicine** and clinical settings, including where HIPAA holds up and where modern inference methods make it vulnerable.

Technical Skills

Languages — Python (NumPy, pandas, scikit-learn), SQL, dbt, R

AI Stack — Omni AI (conversational analytics), Claude Code, Claude + MCP over BigQuery

ML & Statistics — Causal inference, clustering and unsupervised methods, statistical sampling and validation, feature engineering

Cloud, Infrastructure & Tools — Google Cloud Platform (BigQuery, GCS, Cloud Functions, IAM), Fivetran, dbt Cloud, Omni, Looker, Git/GitHub, Jira, Confluence

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