

SCHOOL OF APPLIED DATA

DS-401 Data Science Certificate — Syllabus

Faculty of record: Dr. Amara Okonkwo · 12 weeks · 12 credits · Online with weekly live studio

The certificate trains working learners to clean, model, and explain tabular public data. Students leave with a documented portfolio project, a model card, and a briefing suitable for a non-technical civic partner. Live studio meets Thursdays 6:30–8:00 p.m. Eastern; recordings post within 12 hours.

Learning outcomes

1. Frame a public question that a dataset can actually answer.
2. Build a reproducible Python pipeline with tests for missingness and leakage.
3. Choose evaluation design (holdout, grouped splits) and report uncertainty in plain language.
4. Write a model card and a two-page partner memo.
5. Present findings in 8 minutes without hiding assumptions.

Weekly outline

Wk	Focus	Deliverable
1	Questions, ethics, and data provenance	Project prospectus
2–3	Wrangling, joins, and documentation	Cleaned dataset + codebook
4–5	Exploratory analysis and visualization	Notebook v1
6–7	Supervised baselines and metrics	Model card draft
8	Fairness checks and stakeholder review	Partner memo draft
9–10	Revision studio	Notebook v2
11	Oral briefings	8-minute presentation
12	Portfolio freeze	Final archive in repository

Assessment

Prospectus 10% · Codebook and pipeline 25% · Model card 20% · Partner memo 20% · Oral briefing 15% · Studio citizenship 10%. Late work loses a third of a letter per 24 hours unless an access plan is on file with Student Support.

Required tools

Python 3.12, pandas, scikit-learn, and a Cedarhaven GitLab project. No proprietary cloud account is required. Optional GIS overlay uses QGIS for students also enrolled in GEO-215.

Office hours

Dr. Okonkwo: Tuesdays 11:00–12:30, Hawthorne Library Data Studio, or by video. Teaching fellow hours posted in the course hub.