

FRE 521D: Data Analytics in Climate, Food and Environment

Course Description

This applied course develops practical skills to design and run end to end data pipelines for the climate, food, and environment sectors. Students will source data from files, APIs, and databases; use SQL for reliable access; build reproducible ETL in Python; and create refreshable outputs with clear visual explanations for stakeholders. Emphasis is on business and sustainability use cases, including ESG measurement and reporting.

Course overview

We begin with SQL and relational schemas, move to ETL that includes files and APIs, continue with data wrangling and basic data engineering, and conclude with foundational data analysis.

Scope:

- Establish reliable database access with SQL. Define table contracts and extract the slices required for downstream work.
- Extend access with ETL pulls from files and APIs to enrich climate, food, environment, and ESG datasets.
- Integrate sources in Python, apply tidy merges and reshaping, and record minimal lineage and validation rules.
- Produce analysis-ready datasets that answer a clear decision question and feed concise visual explanations.

Learning outcomes

SQL and schemas

- Read entity relationships, keys, and data types to determine join paths and table intent.
- Write performant queries using joins, CTEs, window summaries, and pivot or unpivot operations.
- Implement simple incremental pulls using timestamps or surrogate keys and document view contracts.

ETL with files and APIs

- Extract from CSV and JSON and from API endpoints after the SQL baseline is in place.
- Handle authentication, parameters, pagination, and rate limits for stable API retrieval.
- Store raw inputs and cleaned outputs separately and note assumptions in a short readme.

Wrangling and data engineering basics

- Apply tidy principles for merging, reshaping, and type control across multiple sources.
- Add lightweight validation checks for ranges, null patterns, and key integrity.
- Record brief lineage notes that link tables and figures to their inputs and transforms.

Data analysis basics

- State the decision question, select measures and groupings, and run descriptive summaries and trend checks.
- Produce small, reproducible tables and charts that address stakeholder questions in climate, food, or ESG contexts.

Applications in climate, food, environment, and ESG

- Build a core SQL view for sector indicators, enrich with API metrics, validate with rules, then present a short analysis that explains a seasonal pattern, a regional anomaly, or a compliance flag.
- Prepare a refreshable slice for a decision brief that lists source tables, API endpoints, key parameters, and the refresh cadence.

Practical skills

- Joins, CTEs, window functions, pivot and unpivot, incremental pulls, and performance-aware filtering in SQL.
- API retrieval with authentication, pagination, throttling controls, and structured saving of raw and cleaned data.
- Python merges and reshape operations with type checks and simple validation under tidy data conventions.
- Short documentation of data contracts, lineage notes, and run instructions suitable for handover.

Tentative Datasets

- Wine Quality Dataset
- ESG Rating Data
- Wildfire Dataset
- Water Quality Data

Assessment Structure

- **Assignments (2) and Labs 35%:** A1 builds a SQL access layer with documented table contracts, joins, and an analysis-ready view. A2 implements an ETL pull from one file source and one API, saves raw and cleaned layers, and includes brief validation notes and run instructions.
- **Group Project and Report 30%:** End to end pipeline for a climate, food, or environment question. Includes a refreshable SQL base, at least one API integration, tidy wrangling with validation checks, a small reproducible figure tied to the decision question, and a short written brief. Submit one repository with clear data lineage and how-to-run steps.
- **Quizzes (3) 30%:** Short quizzes covering SQL and schemas, ETL and API fundamentals, wrangling and validation, and visualization basics.
- **Case Study or Reading 2%:** Concise memo that links a sector dataset to a decision question, states sources and assumptions, and proposes a refresh cadence.
- **Participation 3%:** Active engagement in labs and discussions, peer feedback on pipelines, and timely progress check-ins.

Course Outline

- **Mon Jan 5, 12:30–2:00:** *Introduction to Data Analytics* Course overview, setup of Python, SQL tools, and Git. Principles of tidy data and reproducibility. Short tour of sector datasets used in the course (Wine Quality, ESG Ratings, Wildfire, Water Quality). **Teams formed:** Fri Jan 10.
- **Wed Jan 7, 12:30–2:00:** *Database Basics and SQL Refresher* Schemas, keys, data types, joins, CTEs, and data contracts. Writing queries that produce analysis-ready slices.
- **Mon Jan 12, 12:30–2:00:** *ETL Pipeline I* Extracting from Files and Tables Structured ingest from CSV and JSON, staging raw vs cleaned layers, basic lineage notes.
- **Wed Jan 14, 12:30–2:00:** *ETL Pipeline II APIs and Parameters* API requests, authentication, query parameters, pagination, and rate limits. Saving pulls with clear formats. **A-1 due:** Wed Jan 14 (11:59 pm).
- **Mon Jan 19, 12:30–2:00:** *ETL Tools* Azure Data Studio, BigQuery and API Extractions Using Azure Data Studio for exploration and exports. Building small Python scripts that automate API pulls. **Quiz 1:** Mon Jan 19.
- **Wed Jan 21, 12:30–2:00:** *Python Wrangling I Tidy, Missingness, Validation* Column typing, reshaping, missing-data strategies, lightweight checks for ranges, keys, and nulls.
- **Mon Jan 26, 12:30–2:00:** *Python Wrangling II Merges and Reshaping* Multi-source joins, pivot and unpivot patterns, preparing analysis-ready tables for downstream use.
- **Wed Jan 28, 12:30–2:00:** *Introduction to Data Analysis Basics* Framing a decision question, choosing measures and groupings, descriptive summaries, simple trend checks. **Quiz 2:** Wed Jan 28.
- **Mon Feb 2, 12:30–2:00:** *Data Cleaning I Formats, Types, and Keys* Standardize column names and units, parse dates and times, coerce types, fix encodings, de-duplicate with defined primary keys, normalize categorical values, and write a short data contract that specifies tidy layout and allowed values. **A-2 due:** Mon Feb 2 (11:59 pm).
- **Wed Feb 4, 12:30–2:00:** *Data Cleaning II Missingness, Outliers, and Validation* Profile datasets, treat missing values by mechanism, flag or winsorize outliers, enforce value ranges and business rules, check join cardinality and key integrity, add lightweight reproducible checks, and save cleaned tables with a brief run guide.
- **Mon Feb 9, 12:30–2:00:** *Visualization I* Matplotlib and Plotly Building story telling figures directly from pipeline outputs, small multiples, clear labeling linked to the question. **Quiz 3:** Mon Feb 9.

- **Wed Feb 11, 12:30–2:00:** *Final Presentations* Final presentations and course wrap-up. **Final presentations:** Wed Feb 11.

Real-world application focus across dates: climate risk indicators from weather and wildfire data, food price tracking with CPI and commodity series, ESG screening and reporting from ratings and water quality metrics.

Lab activities across dates: SQL slices with joins and CTEs, API pull scripts with pagination, Python merges and validation, figure-from-pipeline exercises.

Course Topics: Real-World Applications and Questions

Big Questions and Real-World Applications

- How do we turn raw climate, food, and ESG data into refreshable, analysis-ready datasets and concise decision briefs for producers, retailers, lenders, or policymakers?
- Which SQL access contracts and patterns ensure reliable weekly or daily updates, and how do we extend them with API pulls while keeping clear provenance and assumptions?
- How should we structure small ETL pipelines so another analyst can run, validate, and hand over results with confidence and minimal setup?
- What story should our visuals tell about seasonality, regional anomalies, or risk flags, and how do we represent uncertainty and data quality responsibly?

Career Relevance and Competencies

- Roles: data analyst, BI analyst, sustainability analyst, junior data engineer in food and resource sectors.
- Competencies: advanced SQL for access and shaping, ETL from files and APIs with authentication and pagination, Python wrangling with type control and validation, reproducible small-scale data engineering practices, visual storytelling and short written decision briefs.

Connections to FRE domains

- Climate: weather and wildfire series built into a SQL view, API enrichment for recent events, validation checks, and a simple map or time-series narrative for risk screening.
- Food systems: food CPI and commodity price series with a refreshable slice, seasonal indexing, and procurement or pricing insights tied to a clear decision question.
- Environment and ESG: ESG ratings combined with water quality or facility metrics, screening and compliance flags, documented refresh cadence, and audit-ready tables with lineage notes.