BEM 106: Data Science in Economics and Finance
Course Syllabus – Spring Quarter / 2026
Division of the Humanities and Social Sciences
Caltech

Course Instructor

Prof. Pawel Janas (he/him)
Office: Baxter 135
Email: pawel@caltech.edu
TA: TBD
TA Office Hours: TBD
Prof. Office Hours: Tuesdays, 4:00 – 5:00 PM in Baxter 135

Course Description

This course introduces students to the principles and practice of data science in economics and finance. The goal is not only to learn technical tools, but to understand how empirical researchers design, evaluate, and interpret data-driven analyses. The course is structured around the data science pipeline: formulating questions, constructing datasets, representing data, estimating models, evaluating predictions, and translating results into decisions. Along the way we develop a framework for distinguishing prediction from causal inference, understanding heterogeneity in economic data, and evaluating models critically rather than treating them as black boxes. Students will learn how to work with structured and unstructured data, including text, and will gain exposure to modern machine learning methods. The course also introduces large language models and their role in modern empirical workflows, including how they can be used for data construction, labeling, and analysis. The course emphasizes verification, reproducibility, and critical evaluation of model outputs.

Course Goal

By the end of the term, students should be able to design and evaluate empirical data science projects in economics and related fields.

Learning Outcomes

By the end of the course students should be able to:

- Formulate empirical research questions suitable for data-driven analysis
- Distinguish between causal inference and prediction problems
- Understand regression as a tool for both explanation and prediction
- Recognize sources of bias such as omitted variables and measurement error
- Represent economic data through feature engineering and transformations
- Identify heterogeneity and structure within datasets
- Apply clustering and supervised learning methods appropriately
- Evaluate predictive models using appropriate metrics and validation methods

Prerequisites

EC 11 is strongly preferred. Students should have familiarity with statistical programming, mathematical statistics, finance and economics. Experience with applied econometrics and Python programming is strongly recommended.

Required Text: None

Reference Text

A fun and light read is *Mostly Harmless Econometrics* by Angrist and Pischke. A not-so-fun but comprehensive one is *Introductory Econometrics: A Modern Approach* by Wooldridge. Your favorite AI tool is also (probably) a good resource as well. My slides, the notes that you take during lecture, quizzes, and the homework assignments are sufficient to prepare you for the exams.

Course Website: see Canvas

Assessment

Homework Assignments	10%
Attendance quizzes	20%
In-Class Midterm	35%
In-Class Final	35%

A : >93% A- : 93-90% B+ :90 – 87% B :83-87% B- :80-83% C+ : <80%

No Pass/Fail option. Please do not turn in hand-written work. All hand-written work will automatically get a zero.

Attendance and Participation

Class attendance and participation are core tenets of this class. As Caltech students who are, at this point, expert AI prompt engineers, you have the skills to access sophisticated technical solutions to all problems sets (hence the downgrading of homework sets in assessment). However, technical skills are only useful if you know when and how to apply them. Therefore, class time will be dedicated to deepening your understanding of how and why data science is used in economics and finance. I will periodically give unannounced quizzes in class to check for attendance. Make-up exams on different days/times will only be granted under extraordinary circumstances. Please communicate any absences related to family emergencies, religious holidays, conferences, job interviews, or other events that prevent you from attending class as soon as possible.

Wellness Policy

If you find yourself overwhelmed or encountering other personal challenges during the term, please reach out to me so we can develop a plan for you to pursue success in this course in a healthy way. In addition, I encourage you to utilize Caltech's resources (see Canvas).

Students with Documented Disabilities

Students who may need an academic accommodation based on the impact of a disability must initiate the request with Caltech Accessibility Services for Students (CASS). Professional staff will evaluate the request with required documentation, recommend reasonable accommodations, and prepare an Accommodation Letter for faculty dated in the current quarter in which the request is being made. Students should contact CASS as soon as possible, since timely notice is needed to coordinate accommodations. For more information: *https://cass.caltech.edu/*, *cass@caltech.edu*. If you are having difficulties with access or other challenges in the class you think might be related to a disability, but do not yet have a diagnosis, please feel free to reach out to CASS to learn more about resources.

Academic Integrity, Collaboration Policy, and AI

Caltech's Honor Code: "No member of the Caltech community shall take unfair advantage of any other member of the Caltech community." All instances of plagiarism or other academic misconduct will be

referred to the *Board of Control* for undergraduates.

Integrity and programming: in this course, it **IS** acceptable to copy code from generative AI – in fact, **it is the preferred way**. Very rarely will you come across a bug or issue that has not been solved by someone else – instead of re-inventing the wheel, you must learn how to adapt solutions to your own problems, somewhat quickly (as you will be asked to do in industry). I will ask you (pseudo)coding questions on the in-class midterm and final, so if you blindly turn in AI generated homeworks, you will probably do well on the homework portion of the grade, but almost surely fail the in-class exams.

My Status as a "Responsible Employee"

As a faculty member, I am required to notify the Institute's Equity and Title IX Office when I become aware of discrimination, sexual harassment, or sex- or gender-based misconduct involving our community members. If one of my students shares such an experience with me, I can help connect them to support resources but will not be able to keep that information confidential as part of fulfilling my responsibility to make sure my students are offered the opportunity to access information and support by the Institute. For more information, you can email *equity@caltech.edu*, go to *equity.caltech.edu*, or review the Institute's *Sex- and Gender-Based Misconduct Policy*.

If you have experienced such prohibited conduct and want to report it or speak to a confidential resource, consult the *Equity and Title IX Office's webpage on reporting* for guidance.

Course Schedule

Week	Dates	Topics	HW Due
1	3/31& 4/2	Data Science Pipeline Questions, Data, Measurement Vague into Empirical	None
2	4/7 & 4/9	Causal Inference vs. Prediction Potential Outcomes and Identification ATE	4/12
3	4/14 & 4/16	Regressions and CEF Bias and Misspecification OVB	4/19
4	4/21 & 4/23	Representation and Feature Engineering Transformations, Interactions, FEs, Panels PCA	4/26
5	4/28 & 4/30	Heterogeneous Treatment Effects Trees and Random Forests Causal Forests	5/3
6	5/5 & 5/7	Text as Data Beliefs, Sentiment, Embeddings **In-Class Midterm (5/5)**	None
7	5/12 & 5/14	Using LLMs Generative AI Workflow Reproducibility	5/17
8	5/19 & 5/21	Supervised ML Model Evaluation Overfitting	5/24
9	5/26 & 5/28	Decision Making Loss Functions Interpretability	5/31
10	6/2 & 6/4	Modern Data Science System Testing and Feedback **In-Class Final (6/4)**	None

Academic Resources for Students

- **Tutoring:** The undergraduate dean's office provides a free peer tutoring service; If the course isn't listed, students can talk with the dean's office to arrange for a tutor; *https://deans.caltech.edu/*
- **Writing:** The Hixon Writing Center provides professional writing tutors as well as peer tutors, individual and group writing space, and additional resources; *https://writing.caltech.edu*
- **Registrar & FERPA:** The registrar can answer questions about degree progress, privacy of student records, and course enrollment procedures; *https://registrar.caltech.edu*. The website also lists *Option Representatives* for option-specific advising, policies, and information.
- **Library:** Borrow books, retrieve journal articles, receive guidance about research; *https://library.caltech.edu/*
- **Dean of Undergraduate Students**: Wide-ranging assistance addressing issues (academic and other) for undergraduates; *https://deans.caltech.edu*

Additional Resources for Students

- **Student Wellness Center:** Wide variety of health and wellbeing services; *https://wellness.caltech.edu/*
- **Counseling Services:** Free for all students, regardless of insurance plan; *https://counseling.caltech.edu*
- **Occupational Therapy:** Individual sessions and consultations on building healthy habits and routines, time management, planning and organization, and more. Free for all students; *https://ot.caltech.edu*
- **Center for Inclusion and Diversity**: Resources concerning navigating diversity and inclusion, including staff who can speak with students about challenges of harassment and discrimination; *https://diversity.caltech.edu/*
- **Title IX**: Caltech's Title IX Coordinator (*titleix@caltech.edu*) works with students on issues related to sexual harassment, sexual misconduct, and sex discrimination; *https://titleix.caltech.edu/*
- **Caltech Accessibility Services for Students**: The Accessibility Services Specialist works with students with temporary medical conditions, or mental, physical or learning disabilities on accommodation requests and services; *https://cass.caltech.edu*
- **Residential Support**: Resident Associates (RAs) and Residential Life Coordinators (RLCs) are also resources for TAs and students; *https://residentialexperience.caltech.edu/*
- **Career Advising and Experiential Learning:** Provides resources to help students make career decisions and implement career plans; *https://career.caltech.edu/*