New York University Tandon School of Engineering
Computer Science and Engineering
CS-GY 6053 – Foundations of Data Science
Fall 2024
Professor Reeves

To contact professor: dr3218@nyu.edu
370 Jay Street, 8th Floor, Room 840
Office Hours: Wednesday, 11 AM - 1 PM

Course Prerequisites

Students should be comfortable with basic mathematical and statistical skills in addition to having good programming skills; strong knowledge of probability, statistics, and linear algebra will be helpful. The main consideration is that the assignments and project are time-consuming, so be prepared to devote time and attention to this course.

Course Description

This is a semester-long course worth 3 credit hours.

This course offers students a practical, hands-on introduction to the growing field of "Data Science," and will equip them with the fundamental quantitative and computational analytics used. As data-access and data-driven methods become the norm in modern business and research environments, there is a growing demand for individuals who are able to derive meaningful insight from data in a variety of domains. Simultaneously, students who are experienced in these approaches can also bring value to these domains. The emphasis primarily is on understanding fundamental concepts and applications of data science. Our aim rather is to present fundamental techniques within the context of a larger data mining and decision-making process and use data to solve domain-area problems. In other words, this is more of a course about learning to think, rather than learning to code or use a certain package. Through a combination of lectures and programming-based exercises, this course provides an overview of key concepts, skills, and technologies used by data scientists. The course materials, assignments and project will all be prepared using the python programming language and will focus on a Bayesian approach to data analysis.

Course Objectives

By successfully completing this course, you will be able to:

- utilize numerical computing libraries in Python for performing data analysis (e.g., NumPy, Scipy)
- load, process, and analyze data from a variety of sources using Python (e.g. pandas) libraries
- create data visualizations using standard techniques and tools (e.g., arviz, matplotlib, seaborn)
- apply statistical inference to answer questions with data using Python libraries (e.g., PyMC)

- organize code and related analyses using Jupyter Notebooks
- reason and communicate effectively about techniques to harness information and recognize patterns in data
- consider the ethical and moral implications of a world with unprecedented access to data about people and their activities

Course Structure

Grade Breakdown (%)

- Assignments: 60%
- Final Project: 40%

Course Communication

Announcements -

Announcements will be posted on NYU Brightspace on a regular basis. You can locate all class announcements under the *Announcements* tab of our class. It is your responsibility to check the class announcements regularly as they will contain important information about class assignments and other class matters.

Discussion Forum –

You are encouraged to post your questions about the course content in the Ed Discussion Forum (accessible through Brightspace: **Content > Resources > Ed Discussion**). This is an open forum in which you and your classmates are encouraged to answer each other's questions. I endeavor to respond to any posted questions within 24 hours during the week. If you need a more immediate response to your question, I encourage you to attend office hours for assistance. If you have more specific questions or concerns that cannot be handled via the discussion board or during office hours, you can email me.

Netiquette –

Always use professional language (no netspeak/internet slang) in your discussion board posts and emails. Please always be respectful of your classmates even if you disagree with their ideas.

Readings

The textbook for the course is:

McElreath, R. (2020). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan* (2nd ed.). Chapman and Hall/CRC. https://doi.org/10.1201/9780429029608

This book is available in electronic format via NYU's library. Reading will be assigned for each course meeting which will provide exposure to the material that will be covered in lecture along with Python code to accompany the chapter content. The textbook often goes into more detail than what we have time to cover in lecture but is still worthwhile to read prior to lecture.

You can access NYU's library here: http://library.nyu.edu/

Course Requirements

Participation is paramount to your success in this course. Be sure to log into NYU Brightspace multiple times a week, read all announcements, complete all Learning Modules and assignments on time, attend/view weekly lectures, and complete weekly readings.

Homework Assignments – 8 Total, Multiple Due Dates

60% of Total Grade

Course assignments will be distributed using NYU's High-Performance Computing JupyterHub system. Detailed instructions on utilizing JupyterHub for this course will be distributed prior to release of the first assignment. Assignments will allow you to practice the concepts encountered throughout the course. Documented collaboration is allowed though each student must submit their work individually. Copying is strictly forbidden. **The lowest grade you receive on a homework assignment will be dropped.**

Final Project – Proposal Due: Week 11, Project Due: Week 14

40% of Total Grade

The goal of this project is for the student to gain experience in understanding a substantive problem/question, acquiring data relevant to the problem/question, and applying the techniques covered in the course to address the problem/question. The project will be performed in teams. Students will be responsible for selecting a domain area problem (some example problems will be suggested) and finding the appropriate data. The grade for the project will be divided between problem and background research, proposed solutions (short presentation), and implementation. Additional details will be announced later in the semester.

Generative Artificial Intelligence Use

It is important that the work required by the course is yours. You should not use generative AI tools (ChatGPT, Bard, etc.) for any purpose other than idea generation. Use of these tools is considered academic misconduct. No use of generative AI tools (ChatGPT, Bard, etc.) to produce homework solutions is allowed in this course.

Late Work

Late submission is allowed at a 10% penalty, with an additional 10% penalty per 24-hour period.

Moses Center Statement of Disability

If you are a student with a disability who is requesting accommodations, please contact New York University's Moses Center for Student Accessibility (CSA) at 212-998-4980 or mosescsa@nyu.edu. You must be registered with CSA to receive accommodations. Information about the Moses Center can be found at www.nyu.edu/csa. The Moses Center is located at 726 Broadway on the 2nd floor.

Inclusion Statement

The NYU Tandon School values an inclusive and equitable environment for all our students. I hope to foster a sense of community in this class and consider it a place where individuals of all backgrounds, beliefs, ethnicities, national origins, gender identities, sexual orientations, religious and political affiliations, and abilities will be treated with respect. It is my intent that all students' learning needs be addressed both in and out of class, and that the diversity that students bring to this class be viewed as a resource, strength and benefit. If this standard is not being upheld, please feel free to speak with me.

NYU School of Engineering Policies and Procedures on Academic Misconduct *(from the School of Engineering Student Code of Conduct)*

A. Introduction: The School of Engineering encourages academic excellence in an environment that promotes honesty, integrity, and fairness, and students at the School of Engineering are expected to exhibit those qualities in their academic work. It is through the process of submitting their own work and receiving honest feedback on that work that students may progress academically. Any act of academic dishonesty is seen as an attack upon the School and will not be tolerated. Furthermore, those who breach the School's rules on academic integrity will be sanctioned under this Policy. Students are responsible for familiarizing themselves with the School's Policy on Academic Misconduct.

B. Definition: Academic dishonesty may include misrepresentation, deception, dishonesty, or any act of falsification committed by a student to influence a grade or other academic evaluation. Academic dishonesty also includes intentionally damaging the academic work of others or assisting other students in acts of dishonesty. Common examples of academically dishonest behavior include, but are not limited to, the following:

    1. Cheating: intentionally using or attempting to use unauthorized notes, books, electronic media, or electronic communications in an exam; talking with fellow students or looking at another person's work during an exam; submitting work prepared in advance for an in-class examination; having someone take an exam for you or taking an exam for someone else; violating other rules governing the administration of examinations.
    2. Fabrication: including but not limited to, falsifying experimental data and/or citations.
    3. Plagiarism: intentionally or knowingly representing the words or ideas of another as one's own in any academic exercise; failure to attribute direct quotations, paraphrases, or borrowed facts or information.
    4. Unauthorized collaboration: working together on work that was meant to be done individually.
    5. Duplicating work: presenting for grading the same work for more than one project or in more than one class, unless express and prior permission has been received from the course instructor(s) or research adviser involved.

6. Forgery: altering any academic document, including, but not limited to, academic records, admissions materials, or medical excuses.

Access the entire School of Engineering Student Code of Conduct here: engineering.nyu.edu/academics/code-of-conduct