

DS2500 -- Intermediate Programming with Data
Khoury College, Northeastern University
Spring 2022

Profs. Felix Muzny (he/him & they/them) and Laney Strange (she/her)

Instructor emails	laneys@northeastern.edu f.muzny@northeastern.edu
Course web page	http://course.ccs.neu.edu/ds2500
Piazza	https://piazza.com/northeastern/spring2022/ds2500
Gradescope	https://www.gradescope.com/courses/334279 access code WYPV3Y
Lecture Schedule	Sec 1 TF 8:00-9:40am (RI 236) Sec 2 TF 9:50-11:30am (Remote through Feb 5th, then RI 236) Sec 3 TF 1:35-3:15pm (Remote through Feb 5th, then CH 101)
Lab Schedule (DS2501)	Lab Sec 1. M 8:00-9:40am (WVH 212) Lab Sec 2. M 9:50-11:30am (WVH 212) Lab Sec 3. M 11:45am-1:25pm (WVH 212) Lab Sec 4. M 11:45am-1:25pm (WVH 210B) Lab Sec 5. M 1:35-3:15pm (WVH 212) Lab Sec 6. M 1:35-3:15pm (WVH 210B) Lab Sec 7. M 3:25-5:05pm (WVH 212) Lab Sec 8. M 3:25-5:05pm (WVH 210B) Lab Sec 9. M 5:15-6:55pm (WVH 212)

DS2501 is a co-requisite for DS2500; make sure you're signed up for both.

Office Hours

Instructor Office Hours (Online)

Instructor office hours are one-on-one conversations. You can schedule a specific 20-minute slot during office hours. Priority is given to students with an appointment, but you can also just drop in.

- Laney Strange, MW 10am-12pm
 - Online: <https://northeastern.zoom.us/my/laney>
 - Make an appointment: <https://calendly.com/laneystrange/20-minute-office-hours-check-in>
- Felix Muzny
 - Online: <https://northeastern.zoom.us/j/96627964302>
 - Make an appointment: <https://calendly.com/muzny>

TA Office Hours (Online)

- Course TAs will hold regular office hours throughout the week. The website lists the schedule, and we'll post any changes on Piazza.
- Like an in-person office hour, you'll be joining a group of your classmates, with an TA or two hosting the meeting.
- If you need to check-in with someone one-on-one, please set up a time with an instructor.

Recommended Textbook

- Intro to Python for Computer Science and Data Science. Deitel & Deitel. Pearson, 2019. ISBN: 0135404673. Available [free online](#) or [purchase](#).

The textbook is available online via Northeastern's digital library. Relevant chapters will be listed alongside lecture topics on the course website. You do not need to read the textbook ahead of lecture; it's most useful as reference materials or for looking up new examples. Keep it handy when working on the homework or reviewing your lecture notes.

Course Description - DS2500/2501

Provides intermediate to advanced python programming for data science with the aim of preparing students for more advanced courses in data science and to enable practical contributions to software development and data science projects in a commercial setting. Covers object-oriented design patterns using Python, including encapsulation, composition, and inheritance.

Advanced programming skills will cover software architecture, recursion, profiling, unit testing and debugging, lineage and data provenance, using advanced integrated development environments, and software control systems. Through case studies, the course will survey key concepts in data science with an emphasis on machine-learning (classification, clustering, deep learning), data visualization, and natural language processing. Additional assigned readings will survey topics in ethics, model bias, and data

privacy pertinent to today's big data world. Accompanied by DS2501: Lab for DS2500 (1.000 credits) in which students will practice the programming techniques discussed in lecture through hands-on experimentation.

The major topics within the course, and their corresponding textbook chapters, are the following (note that the order in which topics are covered might change):

Text Section(s)	Topics
Ch 1-6	Python review and catch-up
Ch 7	Data types and file manipulation
Ch 10.16	Visualization and time-series analysis
Ch 5.17	Risk analysis and probabilities
Ch 10	Object-oriented programming
Ch 9.1-9.9	Files and exceptions
Ch 11	Computer science thinking
Ch 15.2-15.3	Supervised learning
Ch 15.6-15.7	Unsupervised learning

Evaluation

You will receive one grade that will appear on your transcript for DS2500 and DS2501.

Factor	Number	Weight
Homework	7	10%
Labs	weekly	30%
Project #1 (solo)	1	30%
Project #2 (team)	1	30%
	TOTAL	100%

Letter Grades

Your final grade for DS2500 will use the following breakpoints to convert from letter to number grades.

Letter	Range
A	94-100
A-	90-93
B+	87-89
B	83-86
B-	80-82
C+	77-79
C	73-76
C-	70-72
D	60-69
F	< 60

Attendance Policy

Please do not come to class or lab when you're sick. We'd much rather you stay home and take care of yourself.

Attendance is not required in lecture, unless you are scheduled to give a presentation. If you need to stay home and miss class, there is no need to notify us in advance. Just let us know, and we'll make a recording of one lecture section available to you.

Participation IS required in lab. If you ever need to miss lab, fill out this [Missing Lab](#) form by Sunday night.

Homework Grading Policy

We have weekly homework assignments, related to recent lecture material. Your solution to an assignment will not be traditionally graded. You'll receive full credit for submitting all parts of a

self-reflection. You'll have the option to ask us for feedback on your code, and we'll review and comment on any part(s) that you ask for.

You'll submit two components for each homework:

1. A self-reflection, answering questions about how the assignment went and what you learned.
2. The programming assignment itself, usually .py or .ipynb files.

Your self-reflection will be graded. Answering all questions is sufficient to earn full credit on a homework. We won't look at your programming submission by default, other than your self-reflection. If your self-reflection requests feedback on any part of your work, we'll provide comments in gradescope. You can also come to office hours to go over your approach and ask any clarifying questions.

This policy applies only to weekly homework assignments; projects will be graded traditionally on an A-F scale. If you don't submit a self-reflection, you'll receive a zero on the homework.

Projects

You'll create two projects this semester, based on datasets you find interesting and meaningful, and intended to learn something substantive about that dataset.

The first project is a solo endeavor. You may choose to work with a dataset we've seen in class, as long as you learn something new about it and write your own code to find the answer. You'll present your findings in a lightning talk during lab or lecture. You'll submit your code, but there is no need for a written report. Presentations will be the week before spring break.

The second project is a team effort. You'll choose a new dataset and delve into it with your teammates. Your group will present your findings in a presentation during lab or lecture. You'll submit your code, presentation slides, and a written report summarizing your project. You'll also submit an individual reflection describing your experience in the team. Presentations will be the last 1.5 weeks of classes.

Projects are graded on a traditional A-F scale. They are evaluated based on the quality of your code, the clarity of your presentation, and for project #2 the quality of your written report.

Labs (DS2501)

Participation in labs is required. We use labs to gain hands-on practice with recent lecture material, prepare for upcoming homework assignments, and review previous homework assignments.

Make sure that you attend only the section you are officially registered for. Your instructor will do a brief introduction of relevant material and then you'll have an assignment to complete. Submit your solution on Gradescope at the end of your section.

We expect everyone to attend lab every week. If you are unable to make it, fill out the *missing lab* form

on our course website.

Late/Makeup Policy

- **Homework** - You may submit your homework up to 48 hours late; you'll still receive full credit, but we won't provide feedback.
- **Labs** - Submit your lab work at the end of your section; we'll set aside time during lab so you can submit before you leave. The final deadline for lab submissions is Monday 9pm. No late submissions are accepted, but we drop your lowest lab grade at the end of the semester.

Your attendance in lab is required. Submit your lab work at the end of your section. No late submissions will be accepted. However, we don't expect anyone to attend in-person when they are unwell. If you ever need to miss lab, fill out this [Missing Lab](#) form by Sunday night.

- **Projects and Presentations** - There are two projects this semester, and we expect you to submit them on time and be present for your presentation. However, we don't want anyone to attend in-person when they are unwell. In case of extenuating circumstances, fill out the [Late Project](#) form on our course website at least 24 hours before the submission deadline or your scheduled presentation.

Software

We'll be using Python 3 in this class. Anaconda (<https://www.anaconda.com/>) is your best bet for installing the latest version of Python along with various libraries.

When you install Anaconda, it also comes with the editor Spyder, which we'll use to write and run Python code. Spyder will be our "official" DS2500 editor; if you like and use another editor that's totally fine, but we'll use Spyder in lectures and office hours, and we'll be able to help you out if something goes wrong.

In addition to Spyder, we'll use Jupyter Notebooks in class and labs. Jupyter Notebooks are a great tool to have, but it is not required for any homework.

Communication

The simplest way to get feedback and help from course staff and from your classmates is via Piazza. Piazza is an extension of our classroom discussion, and we expect everyone to behave accordingly. No disrespect, rudeness, or abuse will be tolerated -- towards fellow students or towards the course staff. Piazza will be disabled if we feel it is being misused.

You may not post your code on Piazza, but you can ask, answer, and discuss different things you've tried, what worked and didn't work, and resources you've found.

We'll also use Piazza to post course announcements, so make sure your email settings are turned on!

Email (laneys@northeastern.edu, f.muzny@northeastern.edu) is the best tool for specific questions or concerns about your experience in class or anything sensitive in nature. During the week, we'll respond within 24 hours, but don't expect a response after 6pm. On the weekends we'll be slower to respond, but if you reach out over a weekend you can expect to hear back by Monday morning.

Office hours are the best place for talking through your approach to a homework problem. We're not here to give you answers, of course, but to be your fellow data scientists thinking through a tough problem with you. Expect us to ask more questions than we answer.

Classroom Environment

In our classroom, please ask questions, and answer questions! In programming, we seldom get anything right on the first try. We see how an attempt turned out, and we try again. I like our classroom to reflect that approach as well; so please answer a question that's been posed, even if you're not sure of the answer.

To create and preserve a classroom atmosphere that optimizes teaching and learning, all participants share a responsibility in creating a civil and non-disruptive forum for the discussion of ideas.

Students are expected to conduct themselves at all times in a manner that does not disrupt teaching or learning. Your comments to others should be constructive and free from harassing statements.

Academic Integrity

Because homeworks are not traditionally graded, you are welcome to share your code with classmates. **Do not post code on piazza**, because some of your classmates will prefer to find solutions on their own.

Searching online and looking for ideas is acceptable, as long as (1) you cite any outside sources that you referenced in a comment in your code, and (2) you do not ask TAs or instructors to help you fix code you found online. We'll help you work out problems with your code, not someone else's.

Student Accessibility

If you require support during the course due to a disability please ensure that you are already registered with the University's Disability Center, and contact your course instructors to coordinate any support needed during the course.

Title IX makes it clear that violence and harassment based on sex and gender are Civil Rights offenses subject to the same kinds of accountability and the same kinds of support applied to offenses against other protected categories such as race, national origin, etc. If you or someone you know has been harassed or assaulted, you can find the appropriate resources here: [Title IX](#).