DS2500 -- Intermediate Programming with Data
Khoury College, Northeastern University
Spring 2024
Prof. Laney Strange

Instructor contact	`Call me Laney (she/her)`
	
	`laneys@northeastern.edu`
	https://khoury.northeastern.edu/home/laney/
	
	`Office Hours: T 3:30-6:00pm, F 12-1pm`
	`In-person (ME313) & Zoom (northeastern.zoom.us/my/laney)`
	Make an appointment `or just stop by`

Course web page	http://course.ccs.neu.edu/ds2500
Piazza	https://piazza.com/northeastern/spring2024/ds2500
Gradescope	https://www.gradescope.com/courses/679045
Grading Policies	https://bit.ly/ds2500_guidelines
TA Appreciation	https://bit.ly/ds2500_ta_appreciation
Lecture Schedule (DS2500)	Sec 1 TF 9:50-11:30am (CG 0907) Sec 2 TF 1:35-3:15pm (CG 097)
Lab Schedule (DS2501)	Lab Sec 1. M 8:00-9:40am (WVH 210A) Lab Sec 2. M 8:00-9:40am (WVH 210B) Lab Sec 3. M 9:50-11:30am (WVH 210A) Lab Sec 4. M 9:50-11:30am (WVH 210B) Lab Sec 5. M 11:45am-1:25pm (WVH 210A) Lab Sec 6. M 11:45am-1:25pm (WVH 210B) Lab Sec 7. M 1:35-3:15pm (WVH 210A) Lab Sec 8. M 1:35-3:15pm (WVH 210B) Lab Sec 9. M 3:25-5:05pm (WVH 210A) Lab Sec 10. M 3:25-5:05pm (WVH 210B)

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

Syllabus contents

Office Hours

Laney's office hours begin Thursday, September 7th, and TA office hours begin Monday, September 11th. OH are offered both in-person and online. There are no office hours on university holidays.

Instructor Office Hours (In-person and Zoom)

Laney's office hours are one-on-one conversations. You can schedule a specific 15-minute slot during office hours. Priority is given to students with an appointment, but you can also just drop in.

- Laney's OH. M 2:30-4:30pm, R 9-11am
- In-person location: Meserve 313
- Zoom location: https://northeastern.zoom.us/my/laney
- Schedule time: https://calendly.com/laneystrange/15-minute-office-hours-with-laney?month=2023-09

If my Mon/Thurs times don't work for you, please just send me an email (laneys@northeastern.edu) and we'll set up something else.

TA Office Hours (In-person and Zoom)

TA office hours will be held in person and online (on Zoom -- https://northeastern.zoom.us/j/93466315477). The different in-person locations will be indicated on the OH calendar available on the course website.

For both in-person and online, it'll be important that you come to office hours having already made an attempt on the homework. TAs will be happy to help give you suggestions and provide clarification on concepts. They cannot provide you with answers to problems.

Course Description

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. 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.

Accompanied by DS2501: Lab for DS2500 (1.000 credits) in which students will practice the programming techniques discussed in lecture through hands-on experimentation.

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. You do not need to read the textbook ahead of lecture; it's most useful as reference materials or for looking up new examples.

Course Topics

We will alternate our primary focus each week, switching between programming skills and Data Science techniques. The major topics we expect to cover are as follows

Week	Focus	Topics
1	--	Intro and welcome to DS2500
2	Python	Review of foundational Python programming
3	Data Science	Fundamental statistics in data science
4	Python	Data structures in Python, command-line, data visualization
5	Data Science	Covariance and correlation; scaling and normalizing data
6	Python	Pandas and NumPy
7	Data Science	Supervised learning
8	Python	Web scraping, JSON, APIs
9	--	SPRING BREAK
10	Data Science	Supervised learning cont'd
11	Python	Geopandas and other advanced visualization
12	Data Science	Unsupervised Learning
13	Data Science	Classes & Objects; probabilistic algorithms
14	--	Project Presentations in lab and lecture
15	TBD	

Evaluation

You will receive one grade that will appear on your transcript for DS2500 and DS2501. Your grade will be evaluated based on your homeworks, labs, Python Practice Problems (PPPs), and a team project.

Factor	Number	Weight
Homework	4	40%
Labs	8	20%
Python Practice Problems (2 dropped)	6	10%
Project	1	30%
	TOTAL	**100%**

Letter Grades

Your final grade for DS2500 will use the following breakpoints to convert from letter to number grades. We use natural rounding, so a 95.6 becomes 96 but 95.4 becomes 95. All homeworks are weighted the same, regardless of the specific number of points.

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+	67-69
D	63-66
D-	60-62

F	< 60

Homework Grading Policy

Your homework will be graded on three separate rubrics:

1. **Accuracy**. You'll answer quantitative questions about the dataset on gradescope. These answers are auto-graded. Gradescope can be a little picky, so make sure you don't put extraneous characters or whitespace in your answers -- and double-check the "correct/incorrect" confirmation!
2. **Visualization**. You'll be asked to submit screenshots/downloads of at least one Python plot on every homework. We expect these plots to be labeled, easy to read and understand, and appropriate for the data.
3. **Code Quality.** You'll submit your code as well, which we will review and grade based on its modularity, readability, and reusability. Please see our [DS2500 grading guidelines](#) for more information.

Homeworks will be scored and returned to you, on Gradescope. After getting your homework back, you have the option to file a regrade request. under one of the following categories:

- **Clarity** -- you're not sure why points were taken off, even though you've read the rubric and your grader's comments.
- **Mistake** -- your grader mistakenly took points off.

When filing a regrade request, specify which category the request belongs in. You'll receive a response, and possibly an updated score, from your grader. After that, if you still have concerns or questions about your grade, email laneys@northeastern.edu.

Lab Grading Policy

Labs are scheduled every Monday. You'll have an assignment to work on that is due at 9pm on the same day, though we encourage you to submit at the end of your lab section.

You can work with a partner or alone for labs. If you work with a partner, only one person needs to submit on Gradescope but they need to tag the other person to ensure credit. Labs are auto-graded directly on Gradescope. Because of this, we need you to complete the assignment exactly as specified -- function names, parameters, file names, everything.

You receive full credit for completing and submitting 3 out of 5 problems and passing all tests on gradescope. Feel free to complete all 5 for extra practice. Details of the lab grade:

- 3, 4, or 5 functions pass all tests: 100%
- 2 functions pass all tests: 75%
- 1 function passes all tests: 50%
- 0 functions pass all tests: 0%

PPP Grading Policy

PPPs will be assigned towards the end of most Tuesday lectures. You'll put together code that we'll all use in the following lecture that Friday. We'll have time set aside to do these in class, or you can submit on

your own.

Each PPP is exactly one function, and it will be auto-graded directly on Gradescope. You receive 100% if your PPP passes all tests (it's just one function) and 0% otherwise. We drop 2 of the PPP assignments (out of 6 total).

Project Grading Policy

You'll work with a team to complete a project towards the end of the semester (or you can choose to work alone if you notify us of your preference). You'll choose a new dataset(s) 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.

Projects are evaluated based on the quality of your code (same rubric as the homeworks), the clarity of your presentation, and the quality of your written report.

Communication

Reaching Out to the Teaching Team

The simplest way to get feedback and help from course staff and from your classmates is via Piazza, https://piazza.com/northeastern/spring2024/ds2500

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 don't use Canvas in DS2500. We'll use Piazza to post course announcements and reminders, so make sure your email settings are turned on!

Email (laneys@northeastern.edu) is the best tool for specific questions or concerns about your experience in class or anything sensitive in nature. During the week, I'll respond within 24 hours, but don't expect a response after 8pm. On the weekends I'll be slower to respond, but if you reach out over a weekend you can expect to hear back by Monday morning. Please don't message me on Teams; email is best.

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.

Inclusive Classroom

We believe that diversity and inclusiveness are essential to excellence in academic discourse and innovation. In this class, the perspective of people of all races, ethnicities, gender expressions and gender

identities, religions, sexual orientations, disabilities, socioeconomic backgrounds, and nationalities will be respected and viewed as a resource and benefit throughout the semester. Suggestions to further diversify class materials and assignments are encouraged. If any course meetings conflict with your religious events, please do not hesitate to reach out to Laney to make alternative arrangements.

Questions during lecture
- We invite everyone to raise their hands to ask and answer questions during class, and to engage in discussion with classmates. However, we know it's not always the easiest thing to speak up in a big classroom, or to clarify your thoughts in real time.
- Therefore, you can also ask Laney questions directly via the [DS2500 Lecture Question](). Laney will review these questions during the break and after lecture, and respond during lecture or on Piazza, keeping you anonymous.

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. This includes all ways you interact with classmates and course staff -- in lectures, office hours, Piazza, etc.

Name and Pronoun Usage

As this course includes some discussion, it is vitally important for us to create an educational environment of inclusion and mutual respect. This includes the ability for all students to have their chosen gender pronoun(s) and chosen name affirmed. If the class roster does not align with your name and/or pronouns, please inform Laney of the necessary changes.

Academic Integrity

For labs and the project, you may work with others, and we encourage you to collaborate with your teammates. For those assignments, share freely and discuss widely (but not post code on Ed Discussion, which would disrupt others' work on their own assignments).

For homework and PPPs, your grade reflects an individual effort and your solution must be your own. For these assignments, you can discuss ideas and approaches with classmates, but you may not share code. Specifically:
- You may not post code on Piazza.
- You may not share code with your classmates, either directly or on online forums.
- You may not look at the code written by any classmate.
- You may not look at the code created by any generative AI tool such as ChatGPT.

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.

Violation of this policy has consequences both within the course and through the university's OSCCR office. For a first violation, you will receive a zero on the assignment. For a second violation, you will

receive a failing grading in DS2500. All violations will be reported to OSCCR:
http://www.northeastern.edu/osccr/academic-integrity.

Attendance and Late Policies

Attendance Policy

Attendance is not required in lecture or lab, unless you are scheduled to give a presentation. Please do not come to class or lab when you're sick. We'd much rather you stay home and take care of yourself. If you need to stay home and miss class, there is no need to notify us in advance.

You are still responsible for completing all assigned work. We'll always post notes and code after class, and for some weeks we'll add supplementary videos as well.. These supplementary videos were pre-recorded. **We do not record lectures in DS2500**. If you need to miss class/lab...
- Watch the videos and read the notes posted from the given week.
- Complete and submit the lab and/or PPP for that week.
- Stop by Laney's or a TA's office hours to make sure you're caught up and feeling confident on the material.
- There is no need to notify us about missing class.

These supplementary videos, plus the notes from class, are meant to be helpful for days when you need to miss class, but they will definitely NOT be an identical experience! They should suffice when you miss a class or two due to illness or emergency, but we do not recommend using them as a substitute for regular participation in the in-person lecture.

Late Policies

Homeworks and labs can be submitted late through the policies described below. Your project may not be submitted late. PPPs may not be submitted late, but we drop two of the 6 of them.

Homework Late Policy

Homeworks are generally assigned on Fridays and due two weeks later, on Friday at 9pm, unless otherwise noted.

If you miss a homework deadline, we've set aside two late deadlines where you can submit missing homeworks with no penalty. There will be a separate "Late Submission" assignment on gradescope for every homework that becomes available after the original deadline has passed.
- **HW 1-2**: Late submissions by February 23 at 9pm
- **HW 3-4**: Late submissions by April 16 at 9pm

You may submit your homework under the original deadline OR the late deadline, but not both. Your homework will be graded only once.

This policy exists for times you fall behind on a homework, aren't feeling well, or are busy with classes/life. Apart from these two late deadlines, no late submissions will be accepted.

Lab Late Policy

Labs are scheduled on Mondays and are due at 9pm on the same day. We recommend taking advantage of the in-person lab and working with classmates and TAs, but it is not required. You are still responsible for completing the lab assignment even if you are not present in the lab.

If you miss a lab deadline, we've set aside two late deadlines where you can submit late with no penalty. Because labs are auto-graded, you can simply resubmit them on gradescope up until the late deadline.
- **Labs 1-4**: Late submissions by February 23 at 9pm
- **Labs 5-7**: Late submissions by April 16 at 9pm

Project Late Policy

Your project proposal is due at 9pm on Friday, March 1st. Your final project (slides, code, report; individual reflection) is due on Friday, April 12, and you will present it sometime during class/lab in the last two weeks of the semester. There are no built-in extensions for your project, but please   reach out to me (laneys@northeastern.edu) if something unexpected comes up.

PPP Late Policy

PPPs are generally assigned on Tuesday and due Friday before class, so that we can apply the work you developed to the next lecture. Because of the nature of these assignments, we do not accept late submissions but we drop 2 of them (out of 6 total).

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.

We'll be using Piazza (https://piazza.com/northeastern/spring2024/ds2500) for group conversations and course announcements. Use the chat feature during class to ask questions if you prefer it to raising your hand.

We'll be using gradescope (https://www.gradescope.com/courses/679045) for all labs, homeworks, PPPs, and the project. You should be automatically signed up, but make sure you're all set before the first deadline! Any questions about gradescope, post on Piazza and let us know how we can help.

We'll list deadlines in Canvas, but do not use it otherwise in ds2500.

Students with Disabilities

Students who have disabilities who wish to receive academic services and/or accommodations should visit the Disability Resource Center at 20 Dodge Hall or call (617) 373-2675.

 If you have already done so, please provide your letter from the DRC to me (laneys@northeastern.edu) early in the semester so that we can arrange those accommodations.