**DS2500 -- Intermediate Programming with Data**
**Khoury College, Northeastern University**
**Spring 2025**
**Prof. Laney Strange**

| | |
|---|---|
| Instructor contact | Laney Strange (she/her)<br>(Call me Laney)<br><br>laneys@northeastern.edu<br>https://khoury.northeastern.edu/home/laney/<br><br>Office Hours: M 9-10:30am, R 11am-12:30pm<br>In-person (ME313) & Zoom (northeastern.zoom.us/my/laney)<br>Make an appointment or just stop by |
| Course web page | http://course.ccs.neu.edu/ds2500 |
| Piazza | https://piazza.com/northeastern/spring2025/ds2500 |
| Gradescope | https://www.gradescope.com/courses/915440 |
| Grading Policies | https://bit.ly/3OQ9s88 |
| TA Appreciation | https://bit.ly/ds2500_ta_appreciation |
| Lecture Questions | https://bit.ly/ds2500_lecture_q |
| Lecture Schedule (DS2500) | Sec 1 TF 9:50-11:30am (CH 101)<br>Sec 2 TF 1:35-3:15pm (SN 168)<br>Sec 3 TF 3:25-5:05pm (SN 168) |
| Lab Schedule (DS2501) | Lab Sec 1. M 8:00-9:40am (WVH 210A)<br>Lab Sec 3. M 8:00-9:40am (WVH 210B)<br>Lab Sec 4. M 9:50-11:30am (WVH 210A)<br>Lab Sec 5. M 9:50-11:30am (WVH 210B)<br>Lab Sec 6. M 11:45am-1:25pm (WVH 210A)<br>Lab Sec 7. M 11:45am-1:25pm (WVH 210B)<br>Lab Sec 8. M 1:35-3:15pm (WVH 210A)<br>Lab Sec 9. M 1:35-3:15pm (WVH 210B)<br>Lab Sec 10. M 3:25-5:05pm (WVH 210A)<br>Lab Sec 13. 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 Tuesday, January 7th, and TA office hours begin Monday, January 13th. 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 9-10:30am, R 11am-12:30pm
- In-person location: Meserve 313
- Zoom location: https://northeastern.zoom.us/my/laney
- Schedule time: https://bit.ly/3W4RmmR

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

## TA Office Hours (In-person and OH App)

TA office hours will be held in person and online (on the Khoury Office Hours App via https://admin.khoury.northeastern.edu/). The different in-person locations will be indicated on the DS2500 OH calendar, also 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

Covers object-oriented design patterns using Python, including encapsulation, composition, and inheritance. Advanced programming skills cover software architecture, recursion, profiling, unit testing and debugging, lineage and data provenance, using advanced integrated development environments, and software control systems. Uses case studies to 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 survey topics in ethics, model bias, and data privacy pertinent to today's big data world. Offers students an opportunity to prepare for more advanced courses in data science and to enable practical contributions to software development and data science projects in a commercial setting.

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<br>Review of foundational Python programming |
| 2 | Python | Python review cont'd<br>Object-oriented programming<br>Encapsulation, program design |
| 3 | Data Science | Data science statistics<br>Similarity and distance measures |
| 4 | Python | Data visualization<br>Program and unit testing<br>Debugging |
| 5 | Data Science | Correlation and linear regression<br>Scaling and normalizing data |
| 6 | Python | Data structures and mutability<br>How to give a talk<br>**FRIDAY - EXAM ONE (Feb 14th)** |
| 7 | Data Science | Web scraping, HTML<br>APIs<br>Structured Data - JSON |
| 8 | Python | **Mini-Presentations in Monday lab and Tuesday lecture**<br>**\*\*No Class Friday Feb 28\*\*** |
| 9 | -- | SPRING BREAK |
| 10 | Python | Pandas and NumPy<br>Google Colab |
| 11 | Data Science | Supervised Learning<br>Logistic Regression<br>KNN Classification |

| 12 | Data Science | Prediction Models<br>Simulations and probability distributions<br>Geopandas and advanced visualization |
|---|---|---|
| 13 | Data Science | Unsupervised Learning<br>K-Means Clustering<br>**FRIDAY - EXAM TWO (Apr 4)** |
| 14 | Data Science | Natural Language Processing<br>N-grams and predictor tools |
| 15 | Last Day | Topic TBD |
| Finals Week | | Extra credit opportunity (optional!)<br>Project presentation in-person |

# Evaluation

You will receive one grade that will appear on your transcript for both DS2500 and DS2501. Your grade will be evaluated based on your homeworks, labs, exams, mini-presentation, and project.

| Factor | Number | Weight |
|---|---|---|
| Homework | 5 | 25% |
| Labs | 5 | 20% |
| Exams | 2 | 30% |
| Mini-Presentation | 1 | 5% |
| Project | 1 | 20% |

## 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 | 95-100 |
| A- | 90-94 |

| | |
|---|---|
| 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 |

## Homeworks

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.
- **Correction** -- 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.

## Labs

Labs are scheduled on Mondays. 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%

## Exams

We'll have two exams during the semester. Both are administered in-person during lecture. Please make sure they are on your calendar so you don't miss class that day. Exam dates are:
- Friday, February 14th
- Friday, April 4th

Exams will be on paper. You may bring one 8.5x11-inch cheat sheet with anything written or typed on it, one side only. No other materials will be permitted. You will have the entire 100-minute class period to complete each exam, but they are designed to be shorter. Take your time, answer all questions completely, and double-check your work.

If you have a DAS accommodation related to exams, it is your responsibility to arrange to take the exams in the DAS office. Make sure you set this time up at least a week ahead of the scheduled exams to guarantee the time and space you need.

## Mini-Presentation

The mini-presentation is a team-based, short overview of a dataset using two slides and a maximum of a 5 minutes. You'll submit and present as a team during lab on February 24th or lecture on February 25th. Details are in https://bit.ly/ds2500_miniprez

You'll be graded out of 10 points; 5 for the plot and 5 for the presentation, along the following rubric:
- 5 = plot is interesting and informative; presentation is clear.
- 3 = plot needs improvement, and/or presentation is hard to follow
- 1 = plot is hard to understand, and/or presentation is not useful or informative.

## Semester Project

You'll work with a team to complete a project over the course of the semester (or you can choose to work alone if you notify us of your preference). Details are in https://bit.ly/ds2500_proj

You'll choose a new dataset(s) and delve into it with your teammates. We expect a substantive project that incorporates the contributions of all teammates. We recommend that you work with the same team on your mini-presentations (week of February 24), but this is not required. Three separate components will

be graded (all written material; there is no code submission and no required presentation for the semester project):

1. [Proposal](#) - 10% of project grade.
2. Abstract - 5% of project grade.
3. Report - 85% of project grade.

# 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/spring2025/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, mini-presentations, and the project, you are permitted and expected to 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 Piazza, which would disrupt others' work on their own assignments).

For homework and exams, your grade reflects an individual effort and your solution must be your own. For homework 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, (2) you do not copy code generated by AI, and (3) you do not ask TAs or instructors to help you fix something 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](#).

## Attendance and Late Policies

Please note the key dates where you'll need to be in class:
- **Friday, Feb 14**: Exam one in lecture
- **Monday & Tuesday, Feb 24-25**: Mini presentations in lab and lecture
- **Friday, Apr 4**: Exam two in lecture
- **Date TBD**: During finals week, optional extra credit presentation (in-person only)

We'll use our assigned final-exam slot for the extra credit option; this is usually announced a few weeks into the semester. Please make sure you review the dates above, especially when you're making your travel plans!

## Attendance Policy

Attendance is not required in lecture or lab, except for the dates listed above. 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 homework 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.

### Homework Late Policy

You can submit homeworks up to 48 hours late with no penalty. No other late submissions will be accepted. This policy exists for those times you're having a tough week, are feeling sick, or are falling behind in your work; we won't make any exceptions to this policy.

At the end of the semester, you can resubmit one homework for a new grade. You can review and consider the feedback you receive from us as part of your graded work, and use that to revise and build a better solution. Your second-chance homework may not be submitted late.

If you ask for an extension beyond the 48 hours, I'll remind you that this policy exists for those weeks when you're especially busy, or sick, or have a family issue; no one is entitled to additional extensions except under extraordinary circumstances that go far beyond these common issues. If something very serious is going on, I'm going to be worried about you but won't grant any extensions unless I hear directly from WeCare or a similar source.

It's better to submit something than nothing! Even if your homework is incomplete, if that late deadline is approaching, submit whatever you have. We give partial credit, and it's better to have that than a zero.

## 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. Solutions will be released in between the two late deadlines.
- **Labs 1-3**: Late submissions by February 7 at 9pm
- **Labs 4-5**: Late submissions by March 28 at 9pm

## Mini-Presentation Late Policy

The slides for your mini-presentation are due at 9pm on Sunday, February 23rd. Presentations take place on Monday February 24th and Tuesday February 25th. Laney and TAs will be managing the presentations, so we need to have your slides submitted on time in order for you to present.

If you miss the slide submission deadline, or miss your presentation slot, you can make up most of the points by presenting your final project instead during final-exam week.

## Project Late Policy

Your project proposal is due at 9pm on Friday, February 28. Your final project (report, abstract, and individual reflection) is due on Tuesday, April 15 at 9pm. There are no built-in extensions for your project, but please reach out to me (laneys@northeastern.edu) if an unforeseeable emergency will cause you to miss the deadline.

## Project Presentations for Extra Credit

We'll have an option for in-person project presentations during our official DS2500 final exam slot (the schedule will be announced by the registrar in the first few weeks of the semester). This is an optional chance to raise your overall course grade by up to 3 points. You can use this if...
- you were not able to submit and present your mini-presentation by the deadline; or
- you want an extra-credit opportunity to improve your overall grade in DS2500.

During our final-exam slot, you can present your semester project. All team members who participate in the presentation will be eligible for the extra credit, and the points will be assigned individually.

## Exam Late Policy

We expect everyone to be present for scheduled exams. In the case of an unforeseeable, unavoidable emergency that causes you to miss an exam, reach out to Laney directly. I can't guarantee we'll be able to reschedule, but the earlier you reach out the better.

# Software

We'll be using Python 3 in this class. PyCharm will be our "official" DS2500 editor; if you like and use another editor that's totally fine, but we'll use PyCharm in lectures and office hours, and we'll be able to help you out if something goes wrong. We'll also learn about Google Colab (a cousin of Jupyter Notebook) later in the semester, and encourage its use for your semester project as well.

We'll be using Piazza (https://piazza.com/northeastern/spring2025/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/915440) for all labs, homeworks, and the project. Exams will be administered on paper but will scanned and saved on Gradescope as well. 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.

# Student Services

If you require support during the course due to a disability please ensure that you are already registered with the University's Disability Access Services, and contact Laney 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.