

NORTHEASTERN UNIVERSITY  
**DS 3000: Foundations of Data Science**  
Fall 2024 Course Syllabus (CRNs: 14641, 16053)

**Professor Info**

<b>Instructor:</b>	Dr. Eric Gerber	<b>Office:</b>	ME 331
<b>Email:</b>	<a href="mailto:e.gerber@northeastern.edu">e.gerber@northeastern.edu</a>	<b>Email Prefix:</b>	[DS 3000]:

**Links**

<b>Canvas Page:</b>	<a href="#">Canvas Login</a>		
<b>Gradescope:</b>	<a href="#">Gradescope Link</a>	<b>Piazza:</b>	<a href="#">Piazza Link</a>
<b>OH Link:</b>	<a href="#">OH App Link</a>	<b>OH App Help Guide:</b>	<a href="#">OH App Help Guide</a>

**TL;DR Summary (See following pages for detailed syllabus)**

- The two main websites for you to use are **Canvas** (for course information) and **Gradescope** (for assignment submission).
  - Piazza and Khoury Office Hours websites are your main sources for help.
  - See links to all four sites above.
- You may **always** email the professor to set up meetings outside of regularly scheduled office hours.
- You are expected to have a introductory level grasp of Python before starting the course. If you need to review, there are some review materials on Canvas.
  - All lab and homework assignments will be completed in Jupyter notebooks. The program you use to open and edit the notebooks (Google Colab, VS Code, Cantor, or Jupyter Notebook itself) is up to you, but you will always upload the raw .ipynb file to Gradescope as your submission.
- You must bring your laptop, pen and paper to every class.
- **There is no late homework accepted.** Homework submitted more than 48 hours in advance will receive extra credit.
- **Grade Breakdown:** In-Class Assignments (5%), Labs (10%), Homework (40%), Midterm Exams (10%), Final Project (35%)
- Please see **Page 7** for important statements regarding Academic Integrity, Disability Access Services (DAS) and other student resources, and Title IX information.
- **Finally:** I am here to **support your learning**. If you have any questions about how the rules outlined in this syllabus are meant to do that, please don't hesitate to ask.

**Class Meeting Times/Classrooms:**

- Section 3 (CRN: 14641): MW 2:50 pm – 4:30 pm (**RI 236**)
- Section 5 (CRN: 16053): T 11:45 pm – 1:25 pm, R 2:50 pm – 4:30 pm (**INV 019**)

**Communication:** Please check your email, Piazza, and Canvas regularly. Be aware that while attendance will not be taken, it is the student's responsibility to pay attention to/hear about announcements made in class and there will occasionally be **random in-class assignments**.

**Deadline Day Response Policy:** The professors and TAs will not answer any questions regarding assignments within 6 hours before the deadline of submission (i.e., 6 pm). **Please start your homework early!** We cannot guarantee that we will be monitoring our email/Piazza that late at night; make sure you give yourself enough time to ask for help on any assignment.

**Course Canvas Page:** All sections for each professor have been cross-listed to the same Canvas page. All relevant information and assignments will be posted here, and you can access both Gradescope and Piazza from this page, and also keep track of your grades. Log in to Canvas at <http://canvas.northeastern.edu> and find our class; you should automatically be added to the course.

**Piazza:** Students are strongly encouraged to make use of [Piazza](#) to ask questions and discuss problems. Both the TAs and the professor will be checking Piazza regularly, and students are also encouraged to answer their fellow classmate's questions.

**Office Hours:** To facilitate your learning and engagement, your TAs will be holding virtual office hours. This course will make use of the Khoury Office Hours App, which you will need a Khoury account to access. The schedule for the Office Hours can be found at: <https://officehours.khoury.northeastern.edu/>:

- Log on to the Khoury Office Hours App (link above).
- All TAs currently hosting hours will have set up a Queue, which you can join.
- When it is your turn to be assisted the TA will assist you.

If you want more help understanding how the Khoury Office Hours App works, there is also a [Help Guide](#).

Additionally, the professor is happy to meet students who drop by his office, Meserve Hall 331, on a first-come, first-serve basis. While there is no strict schedule of availability for this, the professor is usually in the office:

- Monday through Thursday 1:30 pm – 2:30 pm
  - Note that occasionally the professor will be out to lunch during these times.

I'm sometimes in the office outside these hours, and you may also **always** email me, or ask after class, to set up a time for us to meet that works best for you.

**Prerequisites:**

- CS 2510 or DS 2500 (strong background in programming assumed)

**Description:**

Data Science (DS) is an interdisciplinary field of inquiry concerned with the study and application of systematically extracting generalizable knowledge from data using mathematics, probability, and statistics, and using this knowledge to draw useful conclusions. This course is an introductory DS class focusing on the foundations of DS.

The course introduces methods and concepts from linear algebra and probability that form a basis for modern machine learning. Emphasizes computational aspects using the Python programming language (the course assumes familiarity with Python). Students work with tensors (in NumPy) and may be tasked with implementing from scratch algorithms central to numerical linear algebra and introductory machine learning. More specifically, this class covers:

- understanding basic theoretical linear algebra as it pertains to machine learning
- working with applied linear algebra in standard numerical computing libraries (e.g. NumPy)
- loading, processing and integrating data from a variety of structured and unstructured sources using Python libraries (e.g. pandas, BeautifulSoup)
- visualizing data using basic techniques and tools (e.g. matplotlib, seaborn, plotly)
- applying introductory concepts in probability, statistics, and machine learning using Python libraries (e.g. scikit-learn)
- using a standard DS tool (e.g. Jupyter Notebook).

**Materials:** The majority of class time will be spent writing and running computer code or working out math problems. Each student should bring their laptop, pen and paper to each class and be prepared to submit in-class assignments, when they occur, at the conclusion of the class.

**Textbook and References:** There is **NO** required textbook for this course. Material and lecture notes will be posted and updated on Canvas every week. If you are interested in having a free textbook as a resource, the below are quite useful:

- *Python Data Science Handbook*, by Jake VanderPlas.  
<https://jakevdp.github.io/PythonDataScienceHandbook/>
- *Linear Algebra*, by Jim Hefferon.  
<https://joshua.smcvt.edu/linearalgebra/>
- *Online Statistics: An Interactive Multimedia Course of Study*, by David Lane.  
<https://onlinestatbook.com/index.html>

**Tentative Course Outline:** There is a tentative schedule (updated as needed) on the Canvas page. This is meant to give a rough idea of the topics covered over the course of the semester, and of tentative due dates, but is by no means set in stone. In general, there are five main parts to the course:

- Part 1: Python and Online Data Collection (~ 5 classes)
- Part 2: Data Cleaning and Basic Data Analysis (~ 4 classes)
- Part 3: Introduction to Linear Algebra (~ 5 classes)
- Part 4: Fundamental Machine Learning Techniques (~ 4 classes)
- Part 5: More Advanced Machine Learning (~ 4 classes)

**Lectures:** You are expected to attend every class period, if able. Be prepared to type and run code in every class and to save your work. There will also be a section of the course that is rather math heavy; please also bring pen and paper to take notes.

**Lecture Notes:** All Jupyter Notebooks and relevant in-class material will be posted to Canvas before each class, and Zoom recordings posted afterwards. These materials are meant to help you review, and catch-up should you miss a class. They are not a replacement for attending class.

**Push-up Promise:** To encourage you to attend and pay attention, I will make a promise to you that has worked out for all parties in the past. For every mistake **YOU** catch me make in lecture over the course of the semester, I will do **ONE** push-up at the end of the last day of class. Every tally will occur only after one of you points out a mistake I made (and after we discuss why/how I made the mistake). This not only gives you an incentive to pay close attention in lecture, but gives me an incentive to limit my mistakes. Don't be afraid to point out my mistakes! We all make them, and it really does help learning!

**Guest Speakers:** My hope is that we will occasionally be able to host a visitor who works in the Data Science industry. This is a tentative plan, which will rely on the availability of the visitors. In the past, we have had guest speakers from companies like Novartis, the Detroit Tigers, and Finra.

**In-class assignments (5%):** Occasionally, at the professor's discretion, there will be random in-class assignments that serve as pseudo attendance checks. Usually, they will be only a single question and should be very simple given you have been attending class regularly and keeping up with material. These may be written out by hand or on the computer, depending on the question. These may **not** be made up, and may only be completed in class on the day they are assigned.

**Labs (10%):** We will occasionally practice concepts with small group work in-class. These labs are meant to be low-stakes, and will be due **at the end of the class day (by 11:59 pm)** via Gradescope submission.

**Homework (40%):** There are four planned homework assignments (worth 10% each) that will typically be assigned and due every three weeks, due **Tuesdays by 11:59 pm**. Assignments will be submitted through Gradescope, where solutions are provided via the rubric. These assignments will have several questions, some of which will be relatively open-ended/project-like. Programming code requires thorough commenting, which must be done independently. Students may submit regrade requests within 4 days of the grades being posted. Students may use Office Hours to receive help on assignments, but under no circumstances will a TA (or the professor) do any part of the assignment for you. If there is evidence that you directly copied answers from **ANY** source, you will receive a score of 0 for the assignment, and further instances of academic misconduct will lead to disciplinary action. Please also see the student guide on Canvas for working with AI tools.

**Early/Late Homework:** There will be no late homework accepted except in **extreme** situations. However, students may receive **5% Extra Credit** if their assignment is submitted **more than 48 hours early: by Sunday at 11:59 pm**.

**Midterm Exams (10%):** There will be two midterm exams, one coding take-home and one in-class math which assess your understanding of the material. Each are low-stakes (5% each), and meant only to give the professors a sense of what is working/not-working for you.

**Final Project (35%):** You will be required to conduct a final group project focusing on finding a large dataset of interest, collecting, cleaning and analyzing the data with appropriate statistical techniques, and interpreting the results in an appropriate manner. The following policies are in place to ensure an equitable share of effort is done by all team members on the project:

- To be allowed to participate in the final group project, you must complete **all** of the assignments leading up to when groups are assigned. This is to ensure that all groups are made up of active students. Any student who has not completed all assignments due by the time of the project groups being formed will receive a final project grade of 0.
- At the completion of the project, each group member will be asked to describe the contribution of all other group members with a **Statement of Contributions**. With this information, the instructor may adjust grades of individual group members, including failing students who have not made meaningful contributions to the final project.
- There will be a presentation aspect of the project where each group member will be required to present on something of substance.
- If there are life circumstances preventing your adequate participation in the group project, please reach out to me as soon as possible. The earlier in the semester we're aware of a situation, the more options we have to put supports in place and remedy it.

More details on the project will be shared after the first two weeks of the semester.

**Total Grading Policy:** In-Class Assignments (5%), Labs (10%), Homework (40%), Midterm Exams (10%), Final Project (35%).

There will be no curve in this class, and final grades will **not** be rounded. Letter grades are assigned according to the thresholds met below (e.g. all grades  $87 \leq X < 90$  receive a B+):

A	A-	B+	B	B-	C+	C	C-	D+	D	D-	F
93	90	87	83	80	77	73	70	67	63	60	0

**Important Tentative Dates (MW/TR Sections):**

Project Assigned .....	September 23
Project Phase I Due .....	October 4
Project Groups Assigned .....	October 7
Coding Take-Home Exam Due .....	October 14
Project Phase II Due .....	October 25
Math In-Class Exam .....	October 28/29
Project Phase III Due .....	November 22
Thanksgiving Break .....	November 27-29
Final Project Due .....	December 2

**Academic Integrity:** Under no circumstances may one student view or share their ungraded homework or quiz with another student. Sharing or viewing another students ungraded work will result in a failing course grade. This does not extend to discussion of concepts or ideas, but prohibits any sharing of personal code. Academic dishonesty is not tolerated and in addition to course failure, all violations will be reported to OSCCR: <http://www.northeastern.edu/osccr/academic-integrity>.

Like every computer scientist, you are encouraged to adapt your own code from code you find online, so long as it was not written for this class in any semester. Doing so requires that you attribute credit to the source:

- a quick url link comment (e.g. [stackoverflow](#)) will suffice.
- please also see the student guide on Canvas for working with AI tools.
  - any response to any problem on any assignment that is found to have been generated by AI (or copied from any other source) will result in a zero for the entire assignment and the student will be immediately reported to OSCCR: <http://www.northeastern.edu/osccr/academic-integrity>.
  - Repeat offenses will result in failure of the course.

**Disability Access Services (DAS):** The office is available to assist students who have a legally documented disability or students who suspect that they may have a disability. If you have a disabling condition that may interfere with your ability to successfully complete this course, please contact Disability Access Services: <https://disabilityaccessservices.sites.northeastern.edu/>.

**Student Resources:** Your health and wellness is more important than any assignment. Please use the resources below if you are struggling, and don't hesitate to ask for my help!

- **Counseling Center (24/7 support):** 877-233-9477 or <https://www.northeastern.edu/uhrs/>

**Title IX:** I am a mandatory reporter under Title IX, which means that I am required to report any and all allegations of discrimination to the Title IX coordinator.

Title IX of the Education Amendments of 1972 protects individuals from sex or gender-based discrimination, including discrimination based on gender-identity, in educational programs and activities that receive federal financial assistance.

Northeastern's Title IX Policy prohibits Prohibited Offenses, which are defined as sexual harassment, sexual assault, relationship or domestic violence, and stalking. The Title IX Policy applies to the entire community,

including male, female, transgender students, faculty and staff.

If you or someone you know has been a survivor of a Prohibited Offense, confidential support and guidance can be found through [University Health and Counseling Services](#) staff and the [Center for Spiritual Dialogue and Service](#) clergy members. By law, those employees are not required to report allegations of sex or gender-based discrimination to the University.

Alleged violations can be reported non-confidentially to the Title IX Coordinator within The Office for Gender Equity and Compliance at [and/or through NUPD](#) (Emergency 617-373-3333; Non-Emergency 617-373-2121). Reporting Prohibited Offenses to NUPD does **NOT** commit the victim/affected party to future legal action.

**Changes to Syllabus:** This course syllabus is intended as a guide. The instructor reserves the right to revise any part of the syllabus during the course. Any changes will be announced well in advance, during class time and/or via email.

Some “Encouragement” (via [xkcd.com/844/](https://xkcd.com/844/) and [xkcd.com/1838/](https://xkcd.com/1838/)):





