This version: September 1, 2024.

Data Science for Everyone
Syllabus

DS-UA 111, Fall 2024

Lectures: Tuesdays & Thursdays 2:00–3:15p
19 West 4th Street, rm. 101
Labs: Fridays (details on Albert)

Prof. Louis Mittel
louis.mittel@nyu.edu
Office hours: Fridays 9–10a
60 Fifth Ave., rm. 214

Prof. Tim G. J. Rudner
tim.rudner@nyu.edu
Office hours: Thursdays, 12-1p (starting on Oct 31)
60 Fifth Ave., rm. 613

Teaching Assistants

Name	Email	Office hours	Location (CDS)
Revant Teotia	rt2741	Fri., 12:30p-2:30p	rm. 763
Yucen Lily Li	yl9959	Mon., 3:00p-5:00p	rm. 763
Wiley Wu	ww1445	Fri., 10:00a-11:00a	rm. 244
Sunny Yang	sy2577	Wed., 2:00p-4:00p	rm. 242

Welcome

Welcome to DS-UA 111, the first course in the sequence for the data science major and minor. This course is intended for students who are interested in pursuing a data science major or minor. Students who are certain they will not be majoring or minoring in data science but would like a one-semester elective course should take DS-UA 300: A Survey in Data Science.

This course **does not** have prerequisites and is geared towards those with no prior exposure to the interdisciplinary field of data science. If you have already successfully completed **both** a programming course and a statistics course, you may be eligible to waive DS-UA 111 and continue with the second class in the sequence (DS-UA 112: Principles of Data Science).

This course packs the foundational components of data science into one semester, covering a lot of ground as data science is the union of statistics and machine learning, two fields with their own deep histories and development.

Please be prepared to work hard. We hope to inspire you to continue your studies in data science and to use your newfound knowledge for good.

Course goals

While the primary purpose of this course is to prepare you for further study in data science, by the end of this course, you should be able to:

- Assess the quality, usefulness, and limitations of a dataset, and evaluate the conclusions and predictions that can be drawn from it.
- Descriptively analyze and explore the features of a dataset starting from only a file and some curiosity.
- Build your own models trained on datasets and make predictions using simple yet powerful model designs.
- Better interpret and understand studies you read about in the news and model-based predictions you encounter in the world.
- Think about decisions in a more rigorous way that takes into account uncertainty.

In the process, you'll learn:

- Python programming, a widely used data science programming language.
- The basics of probability, statistical inference, and testing
- The building blocks of both statistics and machine learning: a model, a dataset, an objective, and a method.
- Key ideas common to both statistics and machine learning including: overfitting versus underfitting, supervised versus unsupervised learning, and measures of model fit and accuracy.

Expectations

We expect you to:

- Attend all lectures and lab sessions.
- Submit all homework on time.

- Complete both exams at the scheduled times.

These are necessary but not sufficient requirements for success in this course. In our experience, students who fully engage with all aspects of this course, beyond just the graded components, tend to achieve the highest grades.

Learning data science, including statistical reasoning and programming, is akin to learning a new language. It demands regular practice and continuous exposure to the material. This course is structured to promote and reward ongoing skills practice and regular concept review.

As with any new language, you will likely encounter periods of frustration and confusion. Persevere through these challenges. With consistent effort, you will soon experience the satisfaction of accomplishment and mastery.

Grading & grading policies

This course provides twice-weekly lectures and a weekly lab section in which you enrolled when you registered for the course. **Students are expected to attend all three**. They are not optional. The information and skills that you need to complete your homework assignments and exams will be provided by the professors and TAs.

Your final grade will be based on:

- **Homeworks (30%)**. Students will have approximately 1-2 weeks to complete each of the five homeworks, which we expect to go out approximately every 1-2 weeks throughout the semester.
- **Midterm exam (30%):** The midterm exam is on **Thursday, October 17, 2024**, in lecture.
- **Final exam (40%):** The cumulative final exam is on **Thursday, December 12, 2024**, in lecture.

The exam times are **firm**. If you miss an exam due to some **unavoidable** illness or other serious circumstance, we will make accommodations – potentially allowing you to sit the exam in the subsequent semester. The term "unavoidable" is key here: We will not issue waivers for voluntary vacations, trips, etc. In addition to not administering make-up exams, we do not plan to provide sample exams. However, we will provide study directions and review sessions.

You must turn in all homework **on time**. We will inform you how to turn in your homework, as it will not be accepted by email. Every day late results in a grade level drop: an A becomes an $A-$, an $A-$ becomes a $B+$, and so on. If you are sick or have some other **unavoidable** problem that interferes with your ability to submit your homework on time, please let us know **before** the deadline.

There is no **fixed** *a priori* curve for final grades, but our distribution is fair and in keeping with other introductory courses at NYU. The best way to do well in this course is to do your work to the best of your ability: doing better will always help your grade.

Academic integrity policy

The University is very clear that students' work is expected to be their own and that plagiarism is not tolerated. The same rules apply here:

1. No assignment on which you receive a grade is collaborative.
2. You may consult with others, but all work handed in must be your own.
3. Do not copy another individual's work, answers, or ideas.
4. Do not allow another individual to copy your work, answers, or ideas.

We will pursue the highest possible punishment for those who violate these policies.

On AI/LLM Tools: Online tools and resources (from Wikipedia to ChatGPT to online message boards) may in some cases help you better understand the course material. Experiment with that at your own discretion. However, you may not use these tools and resources in place of doing homework assignments yourself. You must be able to truthfully attest that the homework you submit is your own work. Copying from an AI chatbot will be treated the same way as copying from any other source. Such violations will result in a zero for the assignment, and repeated offenses may lead to further disciplinary action in accordance with university policies.

Academic accommodations

Academic accommodations are available for student accessibility. Please contact the Moses Center for Student Accessibility (212-998-4980; mosescsa@nyu.edu) for further information. We recommend that students requesting academic accommodations reach out to the Moses Center **as early as possible in the semester** for assistance.

Readings

There are assigned readings for each week of lecture. We strongly recommend completing readings ahead of each week's lectures. All readings are from our course textbook, *Data Science for Everyone: The Book!*. We will post the relevant chapter(s) under Brightspace > Content for each lecture.

Expected Course schedule

- **Week 1**: Introduction to Data Science
- **Week 2**: Causal Inference
- **Week 3**: Generalization and Bias
- **Week 4**: Working With Data
- **Week 5**: Statistical Inference
- **Week 6**: Hypothesis Testing
- **Week 7**: Midterm review and exam
- **Week 8**: Regression
- **Week 9**: Regression
- **Week 10**: Prediction
- **Week 11**: Prediction

- **Week 12**: Machine Learning
- **Week 13**: Machine Learning
- **Week 14**: Final review and exam

The midterm exam is scheduled for **Thursday, October 17, 2024 (in class)**. The final exam is scheduled for **Thursday, December 12, 2024 (in class)**.