

Data Science for Everyone (DS-UA 111)

An innovative remote learning experience

Summer 2020

NYU 6-Week Session I

May 26-July 5, 2020

4 credits

Prof. Andrea Jones-Rooy
cde-undergraduate@nyu.edu

Overview

Data Science for Everyone is the flagship undergraduate course of the NYU Center for Data Science and the first course in the sequences for both the data science major and minor. Now in its third year of instruction, this course has proven extremely popular – the waitlist each semester gets longer and longer. Now is your chance to skip the line and take this course in the summer!

This summer, we are offering this course as a **fully remote, mostly asynchronous** experience! This means you can complete this course from anywhere in the world according to the timing that works best for you. The “lectures” will be delivered in short mini-videos rather than 75-minute zoom sessions that are paired with hands-on exercises to practice what you learn right after you learn it.

The only part that requires a bit of synchronizing is weekly office hours in real time with your TA, as well as homework and exam deadlines, which are fixed. Your TA and professor are both also available for one-on-one and small group appointments.

Provided you are ready to roll up your sleeves and work hard: in just six weeks you will earn the full four credits for the course, equipping you to be a sophisticated consumer and producer of data and data-driven insights – and to hopefully continue with your studies in data science!

This course is truly for **all** students – the only pre-requisite is high school algebra – and we especially welcome students with no prior coding or statistics experience. Welcome to the first day of the rest of your life as a data scientist!

Course timing

Asynchronous segments: Lectures & lab instruction

The content segments of the course, including lectures and instructional parts of labs will be asynchronous, which means you do not have to watch them at a specified time. In addition, rather than appearing as traditional 75-minute videos, the lectures will be in the form of series of short videos (approx. 3 per “lecture”) with hands-on practice exercises in between. Each week there will also be recorded supplementary videos and coding notebooks from your TA that will review and extend the lecture material.

That said, while the course is mostly asynchronous, you will need to keep up with all materials, including having watched all relevant recordings and conducted the related practice prior to joining professor or TA office hours. We will provide a recommended schedule for your planning, but you are welcome to do what works best for you. We also do have fixed deadlines for the four homework assignments and exams, so please use those as guidelines for your own pacing.

Synchronous segments: Office hours, deadlines, and exams

There will be weekly set office hours with the TAs each week. These will be scheduled at a time most convenient for all students given their time zones. We ask that you select one of these slots as your “primary” office hour for you to ask questions and reserve the rest for “observation.” This way, all students can count on their questions being answered each week. You may also schedule appointments with the TAs and faculty member as needed.

Four weekly assignments that reflect the material from the week prior will be due every Monday by 9a NYC time (no homeworks during the weeks of exams).

The exams must be completed on the day they are scheduled during your local time. They will be conducted on Classes, will be open book, and you will have 60 minutes to complete them.

Course outline by week

Overall, the pacing of the course is four lectures and two sections per week. We expect that each week you will spend 300 minutes on lectures & practice materials (approx. 75 minutes per day for 4 days) and 200 minutes on TA recordings & office hours (approx. 100 minutes per day for 2 days), plus approximately 20 hours per week on readings, self-guided practice, homeworks, and exams.

The schedule below is what we recommend, but you are free to complete these lectures whenever you like, provide you are prepared for office hours, homework deadlines, and exams. Note this does not include the TA office hours, which will be determined closer to the start of the semester as we understand students’ time zones.

1. Thinking like a scientist, causality, and evaluating data

- a. May 26: Intro & what is data science?

- b. May 27: Thinking like a scientist
- c. May 28: Causality
- d. May 29: Evaluating data

2. Python programming foundations and working with data

Homework 1 due June 1 (9a ET)

- a. June 1: Intro to Python and what is a computer program?
- b. June 2: Python building blocks: data types & sequences
- c. June 3: Tables and finding data
- d. June 4: Importing and organizing data in Python
- e. June 5: Reflect/catch up (no lecture videos)

3. Descriptive statistics, data visualizations, midterm exam

Homework 2 due June 8 (9a ET)

- a. June 8: Descriptive statistics: concepts & code
- b. June 9: Data visualizations
- c. June 10: Midterm review
- d. June 11: Midterm self-study (no lecture videos)
- e. **June 12: Midterm exam (60 minutes)**

4. Python next steps, correlations, and hypothesis tests

- a. June 15: Functions and loops in Python
- b. June 16: Conditional statements in Python
- c. June 17: Correlations between variables
- d. June 18: Hypothesis testing
- e. June 19: Reflect/catch-up (no lecture videos)

5. Linear regression and prediction

Homework 3 due June 22 (9a ET)

- a. June 22: Linear regression I
- b. June 23: Linear regression II
- c. June 24: Prediction I
- d. June 25: Prediction II
- e. June 26: Reflect/catch up (no lecture videos)

6. Classification, ethics, final exam

Homework 4 due June 26 (9a ET)

- a. June 29: Classification
- b. June 30: Ethics in data science
- c. July 1: Final exam review
- d. July 2: Final exam self-study (no lecture videos)
- e. **July 3: Cumulative final exam (60 minutes)**