

Introduction to Data Science (DS-UA 112)



Center for
Data Science

Term:	Spring 2020
Instructor:	Christopher PolICASTRO
Section Leaders:	Shreyas Chandrakaladharan and Ashwin Bhola
Grader:	Ruoyu Zhu, Peeyush Jain, Sarthak Agarwal and Serkan Karakulak.
Contact:	Public/Private correspondences should be sent through Piazza
Website:	NYU Classes: https://newclasses.nyu.edu/portal JupyterHub: https://dsua-112.rcnyu.org Piazza: https://piazza.com/ Gradescope: https://www.gradescope.com/
Lectures:	Mondays and Wednesdays from 4:55pm-6:10pm at Meyer Hall 4 Washington Place, Room 121
Labs:	Section 002 – Fridays from 9:00pm-9:50pm Section 003 – Thursdays from 3:30pm-4:20pm Section 004 – Thursdays from 4:55pm-5:45pm Section 005 – Fridays from 11am-11:50am Section 006 – Fridays from 12:55pm-1:45pm
Instructor	
Office Hours:	Christopher PolICASTRO Thursday's 11:00-12:00pm (60 Fifth Ave., Room 650 and NYU Classes via Zoom Conferencing)
Section Leader	
Office Hours:	TBD
Grader	
Office Hours:	TBD

Course Description

The course will teach principles and techniques of data science to prepare students in the humanities and sciences to solve problems with data. Students will learn methods of

exploration, prediction and inference to detect patterns, determine unknown information from known information, and quantify uncertainties. These computational and statistical methods will empower students to gain practical insights from data. Experience with programming and modelling throughout the semester will help prepare students for courses about databases, statistics or machine learning—particularly within the data science program. In the undergraduate program for data science, Introduction to Data Science precedes [Causal Inference](#), [Responsible Data Science](#), and [Advanced Topics in Data Science](#).

Course Objectives

The course will teach students to better understand problems with data. Learning outcomes include

- Formulating Questions
 - Sampling
 - Bias
 - Experiments
- Manipulating Data
 - Tables
 - Databases
 - Regular Expressions
- Exploring Patterns
 - Plotting
 - Transforming
 - Dimension Reduction
- Fitting Models
 - Linear Regression
 - Logistic Regression
 - Regression Tree
- Quantifying Uncertainty
 - Hypothesis Testing
 - Confidence Intervals
 - Bootstrap

Throughout the semester, the course will be conducted in Python. However, the course will not focus on the specifics of Python. Instead, students will learn programming skills that should apply to different languages. Please see <https://wp.nyu.edu/idss20/> for more information.

Course Requirements

The instructor will hold lecture twice a week for 75 minutes. Lecture will combine instructional lessons and interactive activities. The section leader will conduct section once a week for 50 minutes. Section will include both discussion and lab. Discussions will give the class the

opportunity to review the lecture guided by questions about the material. Labs will assess students' progress. Working in groups, students will solve problems related to the homework assignments alongside the section leader. Students will be expected to spend time studying outside of class. Grading will be determined by labs, homework, projects, and exams.

Resources

- Data Science
 - [Computational and Inferential Thinking](#) by Adhikari and Denero
 - [Principles and Techniques of Data Science](#) by Nolan, Gonzalez, and Lau
- Programming
 - Python for Data Analysis by McKinney, O'Reilly Media, 2012
 - Python Data Science Handbook by VanderPlas, O'Reilly Media 2010
- Applications
 - [Bit by Bit](#) by Salganik
 - Data Science for Business by Fawcett and Provost, O'Reilly 2010

Course Prerequisites

Familiarity with programming is a prerequisite. Students can meet the prerequisite through [Data Science for Everyone](#), [Introduction to Computer Programming](#) (CSCI-UA 2 or CSCI-UA 3), [Introduction to Computer Science](#) (CSCI-UA 101) or equivalent experience. Please contact [staff](#) with any questions about enrollment.

Course Policies

The grade will be based on assignments, exams, and a project:

- Assignments
 - Students will submit assignments through Gradescope and JupyterHub using nbgrader.
 - *Homework*:
 - 6 assignments combining code and short answers questions.
 - 1 assignment (Homework 0) outlining steps for submission along with formatting conventions
 - *Project*
 - Projects are extended homework assignments. Students will not work in groups.
 - *Labs*:
 - 12 assignments combining code and short answers questions. The lowest two scores will be dropped from the final grade.
- Exams
 - Students will complete pencil-and-paper exams in class.

- *Midterm*: Held the 7th week of the semester. If a student misses the midterm, then the final will be worth 25% (see distribution below).
- *Final*: Held the 16th week of the semester. Time and place to be determined by the registrar.

Surveys

The instructors will ask students to complete three anonymous surveys. Surveys are accessible through links on NYU Classes. Survey 1 will be posted in Week 1 to learn the background and interests of the class. Survey 2 will be posted in Week 4 to gather suggestions about the class. Survey 3 will be posted in Week 9 to follow up on the suggestions and to gauge the midterm exam.

Access

Students will use NYU Classes, Gradescope, Piazza and JupyterHub throughout the semester.

- *NYU Classes*
 - Syllabus, Calendar, Zoom Conference, Link to Piazza, Link to Gradescope, Link to JupyterHub
 - Week 1 to Week 16 agendas containing plan for lecture/section along with references and links to materials
- *Piazza*
 - Public and Private correspondence with students and instructional staff
- *Gradescope*
 - Submission of Homework and Projects. Note that Labs are not submitted through Gradescope.
 - With the exception of mid-semester and final grades, the instructional staff will not maintain grades on NYU Classes.
- *JupyterHub*
 - Students can access JupyterHub with their NYU credentials at <https://dsua-112.rcnyu.org>
 - Lecture slides (posted after class) and demonstrations (posted before class). Section handouts (posted before class) and labs (posted before class). Homework and Projects (posted according to the schedule on NYU Classes Calendar).
 - Please log onto JupyterHub in advance of Lecture and Section to start your server. Students will be logged out after half an hour of idleness. Students will be logged out after three hours of use.

Collaboration

Students can collaborate on homework and labs. However, students are responsible for mentioning their collaborators' contributions in their submission. Homework or labs without acknowledgements violates course policies. With the exception of packages, students should

avoid including duplicating code in their homework, labs, and project. If students duplicate code in their programs, then they must provide comments about the source with attributions.

Late Assignments

For homework and projects, each student gets 5 extension days.

- Extensions are rounded up to the nearest day. For example, 1 minute late means 1 extension day.
- After 5 extension days are used, any homework handed in late will be marked off 20% per day late, rounded up to the nearest number of days.
- No homework will be accepted more than 2 days late.
- Any regrade requests must be submitted through Gradescope within 2 days of release of grades.

Assignments will be **due before 12PM** on the day of the deadline.

Grades

The following weights will be used in the assignment of final grades:

Homework	40%
Labs	15%
Midterm	10%
Final	15%
Projects	20%

Instructional staff will increase a letter grade through demonstration of participation

- Participation in Lecture and Section
- Participation in Instructor, Section Leader and Grader Office Hours
- Participation on Piazza

For example, a grade of B will increase to a grade of B+.

At the end of the semester, the instructional staff will weight your Midterm at 0% and Final at 25% granted the change from Midterm at 10% and Final at 15% will improve your letter grade.

If you have questions about your grades, then please come to grader office hours rather than instructor office hours or section leader office hours

Schedule of Classes

The check the Weekly Agenda along with the Calendar on NYU Classes for the schedule.

Week	Topic
Week 1	Overview
Week 2	Probability and Data Design
Week 3	Tables
Week 4	Visualization
Week 5	Text
Week 6	Databases
Week 7	Midterm
Week 8	Spring Break
Week 9	Models
Week 10	Linear Regression
Week 11	Bias-Variance
Week 12	Logistic Regression
Week 13	Classification
Week 14	Inference
Week 15	Big Data
Week 16	Final

University Policies

Academic Integrity

Work you submit should be your own. Please consult the CAS academic integrity policy for more information: <https://cas.nyu.edu/content/nyu-as/cas/academic-integrity.html> – penalties for violations of academic integrity may include failure of the course, suspension from the University, or even expulsion.

Observances and Sick Days

As a nonsectarian, inclusive institution, NYU policy permits members of any religious group to absent themselves from classes without penalty when required for compliance with their religious obligations. The policy and principles to be followed by students and faculty may be

found here: The University Calendar Policy on Religious Holidays
(<http://www.nyu.edu/about/policies-guidelines-compliance/policies-and-guidelines/university-calendar-policy-on-religious-holidays.html>)

If you are unwell, then please do not attend lecture, section or office hours. Please contact the instructional staff about the circumstances. If the absence impacts your completion of an activity, then the instructional staff will work with you to find an alternative time.

Disability Disclosure Statement

Academic accommodations are available for students with disabilities. The Moses Center website is www.nyu.edu/csd. Please contact the Moses Center for Students with Disabilities (212-998-4980 or mosescsd@nyu.edu) for further information. Students who are requesting academic accommodations are advised to reach out to the Moses Center as early as possible in the semester for assistance.