

# Course Number CSCI-UA 9479

## Course Title Data Management and Analysis

### Instructor Information

- Name:
- Office hours:

### Course Information

- Summer 2023 (4 units)
- Lectures:
- Recitations:
- Class cap size: 20 students

### Prerequisites

CSCI-UA 102 Data Structures or CSCI-SHU 210 Data Structures or CS-UH 1050 Data Structures (or equivalent courses) or CS-UY 1134: Data Structures and Algorithms

Not open to students who have taken Database Design and Implementation (CSCI-UA 60). Students that successfully complete CSCI-UA 479 Data Management and Analysis are not eligible to take CSCI-UA 60 Database Design and Implementation.

### Course Description

Extracting, transforming and analyzing data in myriad formats with traditional relational databases as well as non-relational databases. Data modeling, cloud databases, and API programming. Garnering insights using Python and current data analysis tools and libraries.

### Course Topics

- Python
- data formats and retrieving data:
  - data formats (JSON CSV TSV etc.)
  - retrieving data with Python (web, file system, etc.)

#### Page 1

- working with Python data management packages (pandas, numpy)
- (a taste of) data analysis
  - descriptive statistics
  - linear and logistic regression
  - decision trees and random forests
- data visualization and plotting with Matplotlib
- relational databases using SQL:
  - SQLite

- relational database design
- Entity/Relationship (ER) diagrams
- normalization
- SQL syntax
- basic create/read/update/delete
- JOINS
- aggregate queries
- window functions
- NoSQL:
  - MongoDB (a JSON document database)
    - basic create/read/update/delete
    - querying

## Textbooks

The following books are all available for free online reading through the NYU library's license for Safari Books Online. (Here are some instructions from the NYU library on getting the NYU license to work in O'Reilly's apps for iOS and Android). They can also be purchased in hard copy.

- Data Science from Scratch, Second Edition, by Joel Grus
- Python for Data Analysis, Third Edition, by Wes McKinney
- Hands-On Machine Learning with Scikit-Learn, Keras, and Tensor Flow, 3rd Edition, by Aurelien Geron
- Think Stats, Second Edition, by Allen B. Downey
- Using SQLite by Jay A. Kreibich
- Relational Database Design and Implementation, 4th Edition, by Jan L. Harrington

## Personal Computers and Software

It is highly encouraged that you bring a laptop to class.

I recommend installing Anaconda on your machine. We will be mainly using Google Colab notebooks in class which I will share with you. You can make your own copies of these notebooks and run and edit those copies online. However, you may want to run some notebooks locally on your machine. Anaconda installs JupyterLab on your machine which is what I recommend using for this. It also installs lot of other useful software.

## Grades (Assessment, Measurement and Evaluation)

### Page 2

There will be 1 midterm, 1 final exam, 2 quizzes and 4 assignments.

The homeworks are designed to reinforce the lectures and reading materials. I will propose some questions in class. If you answer or post some answer to 75% of the problems, you will get full credit for participation. Two purposes: to make the class more interactive, and to encourage attendance.

The grade for this course will be determined according to the following formula:

<b>Assignments/Activities</b>	<b>% of Final Grade</b>
Assignments (5)	30%
In-class Participation	5%
Midterm Exam	30%
Final Exam	35%

### **Letter Grades**

Letter grades for the entire course will be assigned as follows:

<b>Letter Grade</b>	<b>Points</b>	<b>Percent</b>
<b>A</b>	4.00	95% and higher
<b>A-</b>	3.67	90% - 94.99%
<b>B+</b>	3.33	87% - 89.99%
<b>B</b>	3.00	83% - 86.99%
<b>B-</b>	2.67	80% - 82.99%
<b>C+</b>	2.33	77% - 79.99%
<b>C</b>	2.00	73% - 76.99%
<b>C-</b>	1.67	70% - 72.99%
<b>D+</b>	1.33	67% - 69.99%
<b>D</b>	1.00	63% - 66.99%
<b>F</b>	.00	62.99% and lower

### **Grading Evaluation Scale**

NYU Paris follows the same grading practices as NYU New York. The following grades may be awarded: A, A-, B+, B, B-, C+, C, C-, D+, D, F. In general, A indicates excellent work, B indicates good

work, C indicates satisfactory work, and D indicates passable work and is the lowest passing grade. F indicates failure. There are some additional grades—P for pass, W for Withdrawal—which are awarded administratively.

## **Grade Dissemination**

Graded tests and materials in this course will be returned individually only by request. You can access your scores at any time using the Grade of NYU Brightspace.

## **Course Policies**

### **Class Attendance and Participation**

Students are expected to attend all scheduled classes unless the instructor explicitly informs the class that other ways of doing the work are acceptable.

No student shall leave a scheduled exercise because of the absence of the instructor until a reasonable time has passed. By tradition and as a matter of courtesy, a student should wait ten minutes before leaving.

### **Assignment Policy**

- Students are allowed to collaborate with other students. You must clearly indicate the collaboration for each problem.
- Searching for solutions is considered plagiarism. If you happen to find the solution from external sources (other than reference and notes), you must give proper citation. • In any case, you have to write your own solutions.

### **Submitting Assignments**

- Students are required to typeset their solutions
- Assignments are usually due at 11:59PM and should be submitted via NYU Brightspace • Late submissions will be accepted up to 24 hours after due date with a penalty 25%

### **Professionalism Policy**

Please attend to all university policy and classroom etiquette procedures. Those not heeding the policies will be asked to leave the classroom immediately so as to not disrupt the learning environment. Please arrive on time, be attentive, and respectful for all class meetings. Students who habitually disturb the class by talking, arriving late or other unprofessional behavior may suffer a reduction in their final class grade.

You are expected to comply with NYU's academic integrity guidelines, as outlined on

<http://www.nyu.edu/about/policies-guidelines-compliance/policies-and-guidelines/academic-integrity-for-students-at-nyu.html>

## Disability Disclosure Statement

Academic accommodations are available for students with disabilities. Please contact the Moses Center for Students with Disabilities (212-998-4980 or [mosescsd@nyu.edu](mailto: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.

NYU is committed to providing equal educational opportunity and participation for students with disabilities. It is NYU Paris's policy that no student with a qualified disability be excluded from participating in any NYU Paris program or activity, denied the benefits of any NYU Paris program or activity, or otherwise subjected to discrimination with regard to any NYU Paris program or activity. Any student who needs a reasonable accommodation based on a qualified disability is required to register with the CSD for assistance. Students can [register online](#) through the Moses Center and can contact the Director of the Academic Resource Center with questions or for assistance.

## Course Schedule (Tentative)

The following schedule is subject to change.

Assignments and due dates will be on Brightspace.

The following schedule is subject to change based on the speed at which we get through various topics.

### Topics and Assignments

Week/Date	Topic/Concept
[Week 1]	Data format, retrieving data
[Week 2]	ER and Relational Model
[Week 3]	PostgreSQL, SQL
[Week 4]	SQL, JOINS, queries

<b>Week/Date</b>	<b>Topic/Concept</b>
[Week 5]	NoSQL: MongoDB
[Week 6]	Optional database and final exam