DS-GA 1007
Programming for Data Science

Professor Jeremy Curuksu
Center for Data Science
New York University
60 Fifth Ave, NY 10011, New York
Email: jeremy.cur@nyu.edu

Syllabus

The class covers the following material.
There may be some changes to best fit the class. All important additional information will be provided accordingly.

- **Introduction to Programming in Python**: Basic data types, sequence types, dictionaries, control flow, statements, assignment semantics, packages, read data, write data, execute code
- **Best Practice Programming and Software Engineering**: Modularization and abstraction, functions, classes and objects, debugging, error handling
- **Program Efficiency**: Execution time, algorithm complexity, iterative and recursive algorithms, searching and sorting algorithms
- **Interacting with Programs**: Python IDEs and editors, distributions/packages, virtual environments, Unix command line interface
- **Array Manipulation for Scientific Computing**: NumPy arrays, broadcasting and indexing, mathematical operations on arrays, linear algebra on matrices
- **Data Visualization**: Graphical visualization with Matplotlib
- **Advanced Data Objects**: Pandas data frames, indexing and selection, Boolean and hierarchical indexing, reshaping, plotting — *Project proposals due*
- **Advanced Data Objects**: Pandas data loading and cleaning, handling missing values, advanced data manipulation, storage and file formats

- **Advanced Data Objects**: Aggregation, group operations, merging, joining, statistical analysis — *Project progress reports due*
- **Advanced Data Objects**: Manipulating, analyzing and vizualizing time series with Pandas
- **Environments for Collaborative Programming**: Version control, Git, GitHub
- **Examples of Industrial Applications** — *Final project due*

Day, Time and Place

DS-GA 1007.001 Lecture

Mondays from 6:45pm to 8:30pm EST

Location: 12 Waverly Place, Room G08

DS-GA 1007.002 Lab

Wednesdays from 7:10pm to 8:00pm EST

Location: 19 University Place, Room 102

Labor Day on Monday, September 5th: In Fall 2022 the first Monday of the semester is on September 5th which is Labor Day and the university is closed. The first time the class will meet will be for the Lab on Wednesday, September 7th.

Teaching Assistants

Section Leaders

- Section 002 - Cora Mao [ym1596@nyu.edu]
- Section 002 - Devarsh Patel [dp3324@nyu.edu]

Graders

- Cora Mao [ym1596@nyu.edu]
- Zhongyan Wang [zhongyan@nyu.edu]
- Pranav Singh [ps4364@nyu.edu]

Office Hours

- Instructor: After class or by appointment
- Cora Mao: Mondays from 10:00am to 11:00am
  CDS Building, Room 244 or over Zoom: https://nyu.zoom.us/j/93328887158
- Devarsh Patel: Thursday from 12:00pm to 1:00pm
  CDS Building, Room 244 or over Zoom: https://nyu.zoom.us/j/91703015803

Lecture Notes, Textbook & Documentation

- The Lectures and Labs Jupyter Notebooks provided are sufficient for this course
- Recommended (but not required) textbook for supplementary reading: "Python Data Science Handbook" by Jake VanderPlas
  (also accessible online at `https://jakevdp.github.io/PythonDataScienceHandbook`)
- Key Python packages used in this course: NumPy, Pandas, and Matplotlib, each has a freely acccessible, online, high-quality and concise documentation, which we recommend you consult regularly during your practices

Grading Policy

Project: 50%; Homework: 50%

- Homework during the whole semester: 50% of the final grade
- Final Project at the end of the semester: 50% of the final grade