

## **CDS DS 110: Introduction to Data Science with Python**

Instructor Name: Dr. Kevin Gold

Course Dates: 9/7-12/12

Office Location: MCS 140G

Course Time & Loc.: 3:35-4:25pm, CAS 313

Contact Information: klgold@bu.edu

Course Credits: 4

Office Hours: Tu 3-4:30, Th 1:30-3

TF: Vinay Gudavalli, gvsahit@bu.edu

### **Course Description:**

DS 110 is the first in a two-course sequence (including DS 210) that builds students' competence in computing techniques central to data science.

In DS 110, students will use Python to explore many of the fundamental computer science concepts and processes used in data science. This begins with core computer science topics such as data structures, development of functions and recursion, and leads to topics including object oriented programming and data visualization.

In terms of Python tools for data science, students will learn how to work with numpy, pandas, and matplotlib to analyze real-world data. (See the more comprehensive list of topics below in this syllabus.)

The sequence of DS 110 and 210 works in concert with the 120-121-122 series: as students develop their expertise in the mathematical foundations covered in 120-121-122, they build their abilities to implement those tools and manage actual datasets in the 110-120 sequence.

### **Hub Learning Outcomes**

#### **Quantitative Reasoning I (QR1) Learning Outcomes:**

Students will be introduced to the Python programming language. Students will learn how to use pure Python to write basic functional and object-oriented programs. Students will also learn about core data science libraries such as numpy and pandas and, by the end of the course, students will be able to use these tools to produce descriptive analyses and visualizations of different datasets.

In parallel to learning Python for data science, students will develop and demonstrate their understanding of core concepts in computer science, and their ability to use them to manipulate different types of data. Students will learn about basic CS topics such as algorithms, big-O notation and algorithm runtime, object oriented programming methods, parallel computing, and networking.

Students will master the fundamentals of programming in Python for data science, including learning how to use common libraries such as numpy, pandas, and matplotlib. Using these tools, students will build basic models of the data to uncover and interpret interesting patterns in real-world data.

Students will learn to communicate the findings of their analyses both verbally and as part of a presentation using summary statistics, tabulations, and visualizations.

### **Teamwork/Collaboration Learning Outcomes:**

Students will form teams of 2-4 to complete a half-semester-long project. Students will be provided with feedback and guidance on teamwork by the course instructor, including how to promote team and individual accountability, set expectations, and how to schedule meetings regularly and make meetings effective.

### **Instructional Format, Course Pedagogy, and Approach to Learning**

The course will be primarily lecture-based, but with additional exercises for the audience designed to engage and check understanding. Homework will play a key role in moving from hypothetical concepts to applied problems. The discussion section will serve as additional review as well as giving the students more personalized feedback. The final project will be a chance to let the students guide their own learning toward a personally meaningful goal.

### **Books and Other Course Materials**

**Recommended:** Deitel & Deitel, *Intro to Python for Computer Science and Data Science*

Recommended readings by day are in the schedule at the end of this syllabus.

### **Courseware**

We'll be using Blackboard for distributing assignments, collecting assignments, grading, and the distribution of course notes: <https://learn.bu.edu>

We will mostly be using Google Colab for assignments and course notes: [colab.research.google.com/](https://colab.research.google.com/)

Students who wish to work offline could download and install Anaconda: <https://www.anaconda.com/>

We'll use Piazza for Q&A: [piazza.com/bu/fall2022/ds110](https://piazza.com/bu/fall2022/ds110)

## Assignments and Grading

Course assignments consist of weekly individual programming assignments, and a half-semester long group project.

The individual assignments are meant to evaluate students' knowledge of the corresponding weekly lecture topics.

The team project is student-driven and consists of two parts: in the first part students work in teams of 2-4 to form an interesting data-science question that can be answered through publicly available data, and descriptive analysis using Python. The students submit their proposal in the middle of the semester. The students incorporate any feedback from the instructor, and upon approval of their proposal begin work on their project. As part of their assessment, students will produce peer evaluations. During the final week of the semester, the students present their projects and submit papers.

Grading:

- Midterm exam (15%)
- Individual assignments (50%, lowest dropped)
- Group project (30%)
- Recitation participation (5%)

## Resources/Support/How to Succeed in This Course:

1. To succeed in this course students should attend all lectures, come to discussion prepared with questions, complete all assignments on time, and discuss problems and material with fellow classmates.
2. Students are welcomed and encouraged to visit office hours.
3. The [Education Resource Center](#) offers free individual and group tutoring.
4. Accommodations for Students with Documented Disabilities: If you are a student with a disability or believe you might have a disability that requires accommodations, please contact the Office for Disability Services (ODS) at (617) 353-3658 or [access@bu.edu](mailto:access@bu.edu) to coordinate any reasonable accommodation requests. ODS is located at 25 Buick Street on the 3rd floor.

## Community of Learning: Class and University Policies

1. **Courtesy expectations.** Students are responsible for supporting a courteous learning environment. Please show respect for other students' questions, and maintain an attentive attitude in class. If you have an advanced question to ask, try to make it accessible to the rest of the class.
2. **Attendance & Absences.** Attending class is optional but highly encouraged; behavior in class does carry some small participation weight. Attending recitation carries some participation credit. However, either can be waived with a reasonable request for accommodation. We will in particular waive for religious holidays; see the University [Policy on Religious Observance](#).

3. **Assignment Completion & Late Work.** Assignments will be submitted as PDF and .ipynb files on Blackboard. After the deadline, assignments can be turned in up to 2 days late for a penalty of 20% of the maximum points, or 4 days late for a penalty of 50% of the maximum points. If solutions have been posted, work can't be turned in late after that.

4. **Academic Conduct Statement**

Students are expected to abide by the guidelines and rules of the Academic Code of Conduct. <https://www.bu.edu/academics/policies/academic-conduct-code/>

### **Outline of Class Meetings: Date, Topic, Readings Due, Assignments Due**

**Homework goes out and is due on *Mondays*.**

**Wednesday 9/7** - Introduction

**Friday 9/9** Python expressions

*Readings: Deitel Ch 2*

**Week of 9/12** - Control flow, lists, tuples; HW1 out; matplotlib and numpy; iteration

*Readings: Ch3.1-3.6 & Ch5.1-5.4; Deitel Ch 7; Ch3.7-3.13*

**Week of 9/19** - More iteration; HW1 due/HW2 out; functions

*Readings: Ch 4*

**Week of 9/26** - more on lists, tuples, and arrays; HW2 due/HW3 out; dictionaries and sets; Pandas dataframes

*Readings: Deitel Ch 6, Deitel Ch 7*

**Week of 10/3** - more on DataFrames, HW3 due/HW4 out, strings, review

*Readings: Deitel Ch 8*

**Week of 10/10** - Files and Exceptions (**meet Tuesday**), HW4 due/HW5 out, object oriented programming

*Readings: Deitel Ch 9, Ch 10*

**Week of 10/17** - data structures, HW5 due/HW6 out, recursion, scikit-learn

*Readings: Deitel 11.1-11.5, 15.1-15.3*

**Week of 10/24** - decision trees, HW6 due/HW7 out, project proposals out, regression

*Readings: Ch12*

**Week of 10/31** - midterm review, HW7 due, **midterm**, text processing

**Week of 11/7** - Scraping the web, proposals due, HW8 out; graphs and centrality, version control

**Week of 11/14** - advanced Pandas, HW8 due, peer assessments out; SQL, visualization with Seaborn

*Readings: Deitel Ch 17.2 (SQL)*

**Week of 11/21** - Complexity, HW9 out, Thanksgiving break

*Readings: Deitel 11.6-11.9*

**Week of 11/28** - Analysis of algorithms, other ways to speed up code, peer assessments due, big data

*Readings: Deitel 11.10-11.14, 17.3-17.6*

**Week of 12/5** - Effective communication, HW9 due, lightning talks

**12/12** - Last lightning talks

*~Last updated 8/11/22~*