

Syllabus: DSCI 510

Principles of Programming for Data Science

(4.0 units)

Instructor: *Gleb Satyukov*

Teaching Assistant: *Geet Shingi and Mia Sultan*

Course Producers: *Tejas Rajurkar and Yash Dani*

Course Graders: *Masooma Naqvi and Qasim Siddiqui*

Data Science Program

Thomas Lord Department of Computer Science

Viterbi School of Engineering

University of Southern California

August, 2023

Last updated: August 22, 2023

Contents

Contents	2
1 Administrative	1
1.1 General information	1
1.1.1 Distance Education Network (DEN)	1
1.2 Description	1
1.2.1 Catalog Description	1
1.2.2 Expanded Course Description	1
1.2.3 Course Objectives	2
1.2.4 Methods of Teaching	2
2 Communication	3
2.1 Office Hours	3
2.1.1 Instructor's office hours:	3
2.1.2 TA's office hours:	3
2.1.3 Help after lectures and lab sessions	3
3 Resources	4
3.1 Textbook	4
3.2 Recordings	4
3.3 Other Resources	4
4 Class Structure and Schedule	5
4.1 Class Structure	5
4.2 Class Schedule	6
5 Grading	7
5.0.1 Grading Scheme	7
5.0.2 Lab Assignments	7
5.0.3 Midterm Exam	8
5.0.4 Final Project	8
6 Academic Conduct	9
6.1 Plagiarism	9
6.2 Use of AI/LLMs	9
7 Support Systems	10

Administrative

1.1 | General information

First day of classes: Tuesday August 22, 2023

For more information please visit: <https://classes.usc.edu/term-20233/classes/dsci/>

Please refer to the following table 1.1 for class times and location:

Table 1.1: Class times and location

Type	Time	Day	Location
Lecture	4:00 - 5:50pm	Tuesday	OHE100D
Lab	4:00 - 5:50pm	Thursday	OHE100D

Please make sure to arrive on time for both lectures and lab sessions. However, if you are late - use the back door to enter the room as to avoid disrupting the class when you come in.

1.1.1 | Distance Education Network (DEN)

For DEN students, our team will make sure to make links to the live video stream available ahead of time. There will be recordings of both lecture and lab sessions. These will be published on the D2L platform.

1.2 | Description

1.2.1 | Catalog Description

Introductory programming course for Non-Computer Science Majors. Learn how to program in Python to search, retrieve, analyze and visualize data found on the Web. Learn to work with large data sets.

1.2.2 | Expanded Course Description

This course is designed to serve as an introduction to computer science concepts and build basic programming skills that are specifically geared towards Data Science. This will form a part of the introductory coursework for the program in Communication Informatics. The main objective of this course is to introduce the fundamental concepts of general computer science and computer

programming, and to give students practical hands-on experience in reading and writing computer programs, developing tools to manipulate large data sets.

The course is designed to be accessible for non-Computer Science major students with little or no programming experience, and emphasizes writing programs that are capable of retrieving and manipulating large amounts of data. The first half of the course focuses on Python as a primary programming language, while the second half of the course covers select advanced topics including data visualization, web scraping, database access, and more.

The course will combine lectures, labs, in-class discussions and problem solving, assigned readings, written homework assignments. There will be a proctored midterm exam and a final project.

1.2.3 | Course Objectives

The objectives of this course are to teach students how to write computer code capable of searching and retrieving data, working with (large) data sets, analyzing data and visualizing the results. Specifically, students that successfully complete this course will achieve the following main objectives:

1. Acquire basic concepts in computer science and computer programming
2. Develop sufficient proficiency in Python programming language to write applications capable of searching, retrieving, manipulating, analyzing, and visualizing data
3. Be able to package, deliver and present a complete Python project using the latest technologies in the computer science community

1.2.4 | Methods of Teaching

The primary teaching methods will be group discussions, case study reviews, and lectures. Students are expected to perform self-directed learning outside of class. This will encompass a considerable amount of time practicing computer programming in Python.

There will be weekly lectures and lab meetings, a midterm exam, lab/homework assignments (from the labs) and a final project.

Students are expected to have access to a computer (preferably a laptop) that can run the Python programming language and a recently updated web browser, and to bring it to class. Beyond that, no special computing facility, hardware or software will be necessary for this course.

Communication

2.1 | Office Hours

Class announcements will be made on our D2L platform - make sure you have the notifications enabled for those announcements as they can be last minute and rather important.

Class discussion board and Q&A will be hosted on **Piazza**, use the following link to sign up:
<https://piazza.com/usc/fall2023/dsci510fall23gleb>

Last but not least, to reach any member of the teaching team - simply click on the names on the title page of this document to send us a direct email. Don't hesitate to reach out if you have any questions, we're here for you!

2.1.1 | Instructor's office hours:

On zoom, anytime throughout the week: <https://zoom.gleb.nl>
Otherwise, send me an email (gleb@isi.edu) in order to schedule a quick meeting.
(please mention **DSCI510** in the subject line for faster processing)

2.1.2 | TA's office hours:

Anytime throughout the week.
Send me an email (shingi@usc.edu) in order to schedule a quick meeting.
(please mention **DSCI510** in the subject line for faster processing)

2.1.3 | Help after lectures and lab sessions

In addition to the online office hours, we can accommodate helping students after lectures and/or lab sessions end. Do not hesitate to come to the front and ask a question after the lecture or lab session ends.

Resources

3.1 | Textbook

We will be using *Python for Everybody* book as our official textbook, which you can access for free here: <https://www.py4e.com/book.php>. This book comes with lots of useful examples, some of which will be used in lecture slides and lab session code tasks.

Realistically, we probably won't be able to cover everything that's in the book. However, we will refer to the specific chapters when we go over specific subjects. In Table 4.1 you will find references to relevant chapters in the book where you can find more information on the subject matter that is discussed that week.

For more experienced folks we would recommend reading *Fluent Python, 2nd Edition* book published by O'Reilly, which you can access for free through [USC Libraries](#).

Go to **Find** -> **Databases** -> **S** -> **Safari Books** and sign in using your USC credentials. From there, search for *Fluent Python* and you should be able to see it at the very top as the first result (*look for a lizard on the cover of the book*).

3.2 | Recordings

Recordings of lectures and lab sessions will be available on our D2L platform.

3.3 | Other Resources

Throughout the course we will be using various other resources, links to which will be shared within the corresponding slides and/or handouts. These additional resources are to help in understanding more complex programming constructs. All of these resources are free and easy to access, an example of such resource would be <https://pythontutor.com/>.

Class Structure and Schedule

4.1 | Class Structure

The class is broken down into two phases, separated by the midterm exam. Phase 1 deals primarily with becoming familiar with computer programming and learning basic programming constructs in Python. The midterm is there to test your knowledge of the programming language and various constructs learned in the process.

Phase 2 begins after the midterm and goes further in-depth on data structures, how to gather data from the Web, how to clean, format and analyze the data, as well as visualizing data for the final project.

In short, students will learn how computer programming works as well as how to write their own programs in python. Students can then take their new skills and apply them on (large) data sets.

The course culminates in the final project, where all aspects of the course come together - from data gathering to analysis and visualization. As such, we don't have a final exam, instead you will be graded on the quality of the final project and its components.

4.2 | Class Schedule

Throughout the course we will roughly adhere to the schedule as shown in table 4.1, please note that this may change slightly as we progress through the semester.

Table 4.1: Class schedule

Week	Topics/Activities	Book Chapters
1 (8/21)	Intro to programs, variables, values, types, expressions and statements	Chapter 1, 2
2 (8/28)	Boolean/logical expressions and conditionals, functions and function parameters, flow of execution	Chapter 3, 4
3 (9/4)	Iterations (while and for loops), strings	Chapter 5, 6
4 (9/11)	Python exceptions, file input/output, lists	Chapter 7, 8
5 (9/18)	Dictionaries, tuples, sets, combinations thereof	Chapter 9, 10
6 (9/25)	Python objects, inheritance, operator overloading, creating multiple/dynamic objects/classes	Chapter 14
7 (10/2)	Function parameters, *args, **kwargs, modules and libraries	Chapter 1 - 10, 14
8 (10/9)	Midterm exam	-
9 (10/16)	HTTP, requests library, DOM, Web scraping with Beautiful Soup	Chapter 12
10 (10/23)	Using web services, XML, JSON, web APIs	Chapter 13
11 (10/30)	Web scraping and an introduction to Pandas, Numpy, and other libraries used in dealing with large sets of data	-
12 (11/6)	Regular expressions and an introduction to various database types	Chapter 11, 15
13 (11/13)	Data Visualization techniques, user interaction and experience	Chapter 16
14 (11/20)	Course review, final discussion	-
15 (11/27)	Work on Final Project	-
16 (12/4)	Final Project Due	-

For the official academic calendar please visit <https://arr.usc.edu/academic-calendar/>

Grading

5.0.1 | Grading Scheme

We will be using the following grading scheme for the course:

Lab Assignments:	30%
Midterm Exam:	30%
Final Project:	40%
Total:	100%

Grades will range from A through F.

The following is a rough breakdown for grading:

Points	Grade
94 - 100	= A
90 - 93	= A-
87 - 89	= B+
84 - 86	= B
80 - 83	= B-
77 - 79	= C+
74 - 76	= C
70 - 73	= C-
67 - 69	= D+
64 - 66	= D
60 - 63	= D-
< 60	= F

Grading for the course will be based on the following components:

5.0.2 | Lab Assignments

Following each lecture, there will be weekly laboratory assignments. These assignments are designed to reinforce the content discussed in lectures. Lab assignments are comprised of various coding tasks that students are expected to submit online (the exact methods are to be announced in the first lecture). Assignments are to be completed individually, during the lab session or taken to-go as homework. Unless explicitly stated otherwise, late submissions are not permitted. There are no lab assignments in the week of the midterm exam.

5.0.3 | Midterm Exam

Halfway through the course there will be a midterm exam to test students understanding of the general computer programming concepts as well as the Python programming language. The midterm exam will be held in person for the on-campus students. For off-campus students the midterm exam will be in-person too, with a proctor approved by the DEN@Viterbi team. For more DEN@Viterbi proctor-related information, please visit:

<https://viterbigrad.usc.edu/technical-support/homework-exams/>

For off-campus students who work or reside in Los Angeles, Orange and Ventura counties, examinations will be taken at USC's main campus on the date and time stated on the course syllabus.

5.0.4 | Final Project

The final project is a culmination of the skills acquired in class and applied to a domain of your interest. This is particularly useful for students from disciplines other than computer science.

For example, if you are interested in housing prices you can search for data pertaining to locations and average prices of homes in an area and compare those prices to crime rates as published by the police or sheriff departments. To get to a visualization of the results you will need to write a script that can gather data, parse and clean the data before analyzing it, and ultimately design a visualization of the results in a way that makes sense and yields new insights. Lab assignments in the second half of the class will be geared towards the final project and you may re-use your code in your final project - therefore it would behave you to start thinking of a subject for your final project as early as possible.

Academic Conduct

6.1 | Plagiarism

Plagiarism – presenting someone else’s ideas as your own, either verbatim or recast in your own words – is a serious academic offense with serious consequences. Please familiarize yourself with the discussion of plagiarism in *SCampus* in Section 11, *Behavior Violating University Standards* <https://policy.usc.edu/wp-content/uploads/2021/04/SCampus-Part-B.pdf>. Other forms of academic dishonesty are equally unacceptable. See additional information in *SCampus* and university policies on scientific misconduct:

<https://policy.usc.edu/research-and-scholarship-misconduct/>.

6.2 | Use of AI/LLMs

Use of AI and specifically Large Language Models (LLMs) is *allowed*. However, it is only allowed as a tool to assist in learning. That is to say, that you may use AI models such as **ChatGPT** or **Claude 2** to help understand the assignments, to ask generic questions about programming in Python and help you by generating code samples that could be of use to explain how certain programming constructs work. Submitting assignments completely generated by AI is **strictly prohibited** and when discovered will be awarded 0 points for the assignment. We will be utilizing additional software to check for code generated by an AI.

Support Systems

Counseling and Mental Health - (213) 740-9355 – 24/7 on call

<https://sites.usc.edu/counselingandmentalhealth/>

Free and confidential mental health treatment for students, including short-term psychotherapy, group counseling, stress fitness workshops, and crisis intervention.

National Suicide Prevention Lifeline - 1 (800) 273-8255 – 24/7 on call

<https://988lifeline.org/>

Free and confidential emotional support to people in suicidal crisis or emotional distress 24 hours a day, 7 days a week.

Relationship and Sexual Violence Prevention and Services (RSVP) - (213) 740-9355 (WELL) – 24/7 on call

<https://sites.usc.edu/clientservices/>

Free and confidential therapy services, workshops, and training for situations related to gender-based harm.

Office of Equity and Diversity (OED)- (213) 740-5086 | Title IX – (213) 821-8298

<https://eotix.usc.edu/>

Information about how to get help or help someone affected by harassment or discrimination, rights of protected classes, reporting options, and additional resources for students, faculty, staff, visitors, and applicants. The university prohibits discrimination or harassment based on the following protected characteristics: race, color, national origin, ancestry, religion, sex, gender, gender identity, gender expression, sexual orientation, age, physical disability, medical condition, mental disability, marital status, pregnancy, veteran status, genetic information, and any other characteristic which may be specified in applicable laws and governmental regulations. The university also prohibits sexual assault, non-consensual sexual contact, sexual misconduct, intimate partner violence, stalking, malicious dissuasion, retaliation, and violation of interim measures.

Reporting Incidents of Bias or Harassment - (213) 740-5086 or (213) 821-8298

<https://campussupport.usc.edu/trojans-care-4-trojans/>

Avenue to report incidents of bias, hate crimes, and microaggressions to the Office of Equity and Diversity | Title IX for appropriate investigation, supportive measures, and response.

Office of Student Accessibility Services (OSAS) - (213) 740-0776

<https://osas.usc.edu/>

Support and accommodations for students with disabilities. Services include assistance in providing readers/note-takers/interpreters, special accommodations for test taking needs, assistance with architectural barriers, assistive technology, and support for individual needs.

USC Support and Advocacy - (213) 821-4710

<https://campussupport.usc.edu/>

Assists students and families in resolving complex personal, financial, and academic issues adversely affecting their success as a student.

Diversity at USC - (213) 740-2101

<https://diversity.usc.edu/>

Information on events, programs and training, the Provost's Diversity and Inclusion Council, Diversity Liaisons for each academic school, chronology, participation, and various resources for students.

USC Emergency - UPC: (213) 740-4321, HSC: (323) 442-1000 – 24/7 on call

<https://www.usc.edu/emergency/>

Emergency assistance and avenue to report a crime. Latest updates regarding safety, including ways in which instruction will be continued if an officially declared emergency makes travel to campus infeasible.

USC Department of Public Safety - UPC: (213) 740-6000; HSC: (323) 442-1200 – 24/7 on call

<https://dps.usc.edu/>

Non-emergency assistance and safety information.

USC Viterbi
School of Engineering

Copyright © 2023 University of Southern California