USC VITERBI SCHOOL OF ENGINEERING
DATA SCIENCE PROGRAM

DSCI 510: Principles of Programming for Data Science (4 units)
Spring 2023

Instructor:
Prof. Jose-Luis Ambite   ambite@isi.edu

Course Producers/Graders:
Gursimar Kaur   gursimar@usc.edu
Sanjana Parakh   svparakh@usc.edu
Samuel Sudheer Vara   svara@usc.edu
Swarnita Venkatraman   swarnita@usc.edu

When emailing, please put "DSCI510" in the subject line

Lecture
Day: Wednesday
Time: 4:00pm–5:50pm
Room: RTH 109

Lab
Day: Thursday
Time: 4:00pm–5:50pm
Room: RTH 109

Platform: DEN D2L (not blackboard!): https://courses.uscden.net/d2l/home/25488
Zoom links available in D2L class page.

Textbook (on the web)
Python for Everybody: Exploring Data in Python 3, by Charles R. Severance, plus other materials
You may also wish to consult Think Python (2nd edition), by Allen B. Downey

Instructor's office hours: After class, or by appointment.
Although this is not required, students are advised to make appointments with the instructor ahead of time in any event and be specific with the subject matter to be discussed. Students should also be prepared for their appointment by bringing all applicable materials and information.

Catalogue Description:
Introductory programming course for non-Computer Science majors. Programming in Python for retrieving, searching, and analyzing data from the Web. Learning to manipulate large data sets.

Expanded Course Description:
This course is designed to serve as an introduction to computer science concepts and basic programming skills that are specifically geared toward Informatics and Data Science. The main objective of this course is to introduce the fundamental concepts behind general computer science and programming, and to give students practical hands-on experience reading and writing computer programs, in order to give them the tools to manipulate large data sets.

The course is designed to be accessible to non-Computer Science major students with little or no programming experience, and emphasizes writing programs that are capable of retrieving and

manipulating large amount of data. The first half of the course focuses on Python as a first programming language, while the second half of the course covers selected advanced topics including web scraping, database access/SQL, data manipulation, data visualization, and more.

The course will combine lectures, labs, in-class discussion and problem-solving, readings, written homework assignments, a mid-term exam, and a final project.

Course Objectives:

The objective of this course is to train students to write computer code capable of manipulating large data sets. Specifically, students successfully completing this course will achieve two main objectives:

1. Acquire basic concepts in computer science and programming.
2. Develop sufficient proficiency in Python to write applications capable of retrieving, searching, manipulating, analyzing, and displaying data.

Methods of Teaching:

The primary teaching methods will be discussion, case studies, and lectures. Students are expected to perform directed self-learning outside of class which encompasses, among other things, a considerable amount of programming practice.

There will a midterm exam. There will ***not*** be a final exam, but there will be a final project. There will be weekly lab meetings and short programming assignments.

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

Grading Scheme (still subject to change):

Weekly Lab Assignments: 42% (14*3)
Mid-Term: 25%
Final Project 33% (1st Deliverable: 3%; 2nd Deliverable: 10%; 3rd Deliverable: 20%)
Total 100%

Grades will range from A through F. The following is the rough breakdown for grading:

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

Grading for the course will be based on four major components:

1. LAB ASSIGMENTS – The labs will reinforce the content discussed in the lecture. During the labs some programming problems will be solved. Then, some similar problems will be assigned to be solved either during the lab or in the next couple of days. There will be laboratory assignments given every week. These assignments are to be completed individually.

2. MID-TERM EXAM – A written exam, roughly halfway through the term, covering material taught up to that point, which is expected to be core Python.

3. FINAL PROJECT – A final project that builds on the techniques taught in class. The final project will be a programming assignment that addresses a data science task. The project should integrate and analyze three data sources from the web. There will be three deliverables: 1) a brief description of the proposed project; 2) code and data scraped from at least one source; and 3) final code and a short 4-page paper on the project and its analysis. Further details on the project will be provided.

Class Communication:
USC DEN Desire2Learn (D2L) platform and Zoom. Classes will be recorded.

Books and Readings:
All books, papers or reports will be available to students via the web. A hardcopy version of the textbook can be purchased if students wish to do so. It is not required.

Required Reading:
Python for Everybody: Exploring Data in Python 3, by Charles R. Severance. An open textbook, available online. Potentially other texts, videos and resources from the internet.

Class Structure & Schedule:
Class sequence, dates, topics and guest speakers are subject to change as the semester proceeds. Any revisions will be noted and announced in class in advance.

Week	Topics/Activities	Book Chapters
1 (1/11)	Intro to programs, variables, types, expressions.	1, 2
2 (1/18)	Conditional execution. Functions.	3, 4
3 (1/25)	Iteration. Strings.	5, 6
4 (2/1)	Files. Exceptions. Lists.	7,8
5 (2/8)	Dictionaries, Tuples, Sets.	9, 10
6 (2/15)	Object-Oriented Programming. Python objects, inheritance, operator overloading, creating multiple/dynamic objects.	14
7 (2/22)	Function Params: *args, **kargs. Scope, Modules, Packages, Libraries.	14
8 (3/1))	MIDTERM.	
9 (3/8)	Networked Programs. HTML DOM, requests library, web scraping with Beautiful Soup.	12
10 (3/11)	SPRING BREAK	
11 (3/22)	XML, JSON, web APIs. *Project Proposal due March 24*	13
12 (3/29)	Regular Expressions.	15
13 (4/5)	Intro to databases, DB design and normalization, SQL queries.	15
14 (4/12)	Data manipulation with Pandas and Numpy. *Project Submission 2, due April 14*	
15 (4/19)	Data visualization with Matplotlib and Seaborn.	
16 (4/26)	More advanced Python, Databases, SQL, OR-Mapping *Final Project due May 2*	11

Statement on Academic Conduct and Support Systems

Academic Conduct

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://scampus.usc.edu/1100-behavior-violating-university-standards-and-appropriate-sanctions. Other forms of academic dishonesty are equally unacceptable. See additional information in *SCampus* and university policies on scientific misconduct, http://policy.usc.edu/scientific-misconduct.

Discrimination, sexual assault, and harassment are not tolerated by the university. You are encouraged to report any incidents to the *Office of Equity and Diversity* http://equity.usc.edu or to the *Department of Public Safety* http://capsnet.usc.edu/department/department-public-safety/online-forms/contact-us. This is important for the safety of the whole USC community. Another member of the university community – such as a friend, classmate, advisor, or faculty member – can help initiate the report, or can initiate the report on behalf of another person. *The Center for Women and Men* http://www.usc.edu/student-affairs/cwm/ provides 24/7 confidential support, and the sexual assault resource center webpage http://sarc.usc.edu describes reporting options and other resources.

Support Systems

A number of USC's schools provide support for students who need help with scholarly writing. Check with your advisor or program staff to find out more. Students whose primary language is not English should check with the *American Language Institute* http://dornsife.usc.edu/ali, which sponsors courses and workshops specifically for international graduate students. *The Office of Disability Services and Programs* http://sait.usc.edu/academicsupport/centerprograms/dsp/home_index.html provides certification for students with disabilities and helps arrange the relevant accommodations. If an officially declared emergency makes travel to campus infeasible, *USC Emergency Information http://emergency.usc.edu* will provide safety and other updates, including ways in which instruction will be continued by means of blackboard, teleconferencing, and other technology.