New York City College of Technology / CUNY CityTech
Department of Computer Systems Technology (CST)

CST2312–HD32
INFORMATION AND DATA MANAGEMENT I

4 hours – 3 credits

Section:	Tues/Thurs 8:00 a.m. - 9:40 a.m. ET/USA
Online:	Hybrid (class is in person unless otherwise posted)
In-person:	When In-person: **Namm N-0920**
Instructor**:**	Professor Shivani Sheth
Instructor Email	shivani.sheth64@citytech.cuny.edu
Office Hours:	Fridays, 4:00-6:00. or as scheduled.
Office Location:	https://calendly.com/profshivanisheth/office-hours

Course Description

This course provides students with the introduction to the necessary informatics and intellectual tools to become efficient and effective information users. The course covers topics related to the digital infrastructure, acquisition, organization, management, and curation of data. The course is structured around the Python tools for regular expression analysis, accessing data sources (crawling, Web APIs), analysis of structured data. At the end of the class the students complete a project to demonstrate the mastery of the technical topics discussed in class with an application to their domain of interest.

Course Objectives

Upon successful completion of this course, students should be able to:

1. Demonstrate knowledge of regular expressions
2. Demonstrate knowledge of web APIs
3. Demonstrate the skills of processing information downloaded from Internet
4. Demonstrate knowledge of web crawling

General Education Outcomes:

- **SKILLS/Inquiry/Analysis:** Students will employ scientific reasoning and logical thinking.
- **SKILLS/Communication:** Students will communicate in diverse settings and groups, using written (both reading and writing), oral (both speaking and listening), and visual means
- **VALUES, ETHICS, RELATIONSHIPS / Professional/Personal Development:** Students will work with teams, including those of diverse composition. Build consensus. Respect and use creativity.

Prerequisite

CST 1101 Problem Solving with Computer Programming.

Course Outline

Week	Topics	Assignments and Readings
Thurs, Jan 25		
Tues, Jan 30	Introduction Introduction to IPython and Collab Notebooks	Send a screenshot of your set up environment Assignment 0 out
Thurs, Feb 1	Review of Python: variables, data types, statements, expressions, functions	Assignment 0 DUE Assignment 1 out
Tues, Feb 6	Review of Python: Strings	
Thurs, Feb 8	Review of Python: Loops and lists	Quiz 0: Variables, Statements, expressions, Strings
Tues, Feb 13	Overall Python review: Functions, testing, running code	
Thurs Feb 15	Introduction to tuples, dictionaries and sets	Assignment 1 DUE Quiz 1: Lists, loops, functions Assignment 2 out
Tues, Feb 20	Reviewing tuples, dictionaries and sets	
Thurs Feb 22	NO CLASS: CLASSES FOLLOW A MONDAY SCHEDULE	Assignment 2 DUE Quiz 2: Advanced data types Assignment 3 out
Tues, Feb 27	Introduction to Regular Expressions	
Thurs, Feb 29	Regular Expressions	Assignment 3 DUE
Tues,	Columbus Day: no class	

Mar 5		
Thurs, Mar 7	Midterm Review	Quiz 3: Regular Expressions Assignment 4: Practice Midterm Exam
Tues, Mar 12	Midterm Exam: Lists, Sets, Dictionaries, Tuples, Regex, Conditionals, Functions	
Thurs, Mar 14	Working with Files	
Tues, Mar 19	File IO / using CSVs Setting up numpy and pandas How to use a library	
Thurs, Mar 21	Numpy and Pandas: Dataframe and data series	Assignment 5 out
Tues, Mar 26	Using numpy and pandas in real life	
Thurs, Mar 28	Introduction to Web APIs Crawling, scraping	Assignment 5 DUE Quiz 4: File IO, Numpy, Pandas
Tues, Apr 2	More Web scraping, crawling Discussing final projects	Come up with a few final project ideas
Thurs, Apr 4	Connecting it all together: Using Regex to get data from a website and importing it into pandas	Assignment 6: Web Scraping
Tues, Apr 9	Connecting it all together: Getting data into pandas and analyzing it	Submit final project proposal
Thurs, Apr 11	Python review: What data structure to pick / Time complexity	Assignment 6 DUE
Tues, Apr 16	Python Libraries review, debugging, web scraping review	Submit scaffolding for final project
Thurs, Apr 18	No Class: Thanksgiving	Final Project
	SPRING BREAK	
Thurs,	Class time to work on final projects	Final Project

May 2		
Tues, May 7	Class time to work on final projects	Final Project / study for final
Thurs, May 9	Class Presentations, Final Review	study for final
Tues, May 14	Class Presentations, Final Review	Study for final
Thurs, May 16	Final Presentation	
Tues, May 21	Final Exam	

Textbooks, Readings, and Resources

- Textbook:
  - Python for Everybody: Exploring Information by Charles R. Severance
  - https://www.py4e.com/lessons
- IDE platform:
  - Google Colab at https://colab.research.google.com/
  - IDLE Python at python.org/downloads/
- Python Data Analysis Library: http://pandas.pydata.org/
- Python package for scientific computing: http://www.numpy.org/

S5. Anaconda Data Science ecosystem: https://www.continuum.io/

WEB PLATFORM – ONLINE CLASSROOM AND LEARNING ENVIRONMENT

W1. CUNY's Blackboard environment online will be used for classes (Collaborate Ultra), content, and discussions at https://bbhosted.cuny.edu/ (use your CUNY login)

Evaluation and Grading

EXAMS + QUIZZES	**35%**
Midterm Exam	15%
Final Exam	15%
Quizzes	5%
ASSIGNMENTS	**60%**
Homework	30%
Final Project	30%
PARTICIPATION	
Class Participation	5%
	=====
Total	**100%**

Grade System

Letter Grade	A	A-	B+	B	B-	C+	C	D	F
Numerical Grade	93-100	90-92.9	87-89.9	83-86.9	80-82.9	77-79.9	70-76.9	60-69.9	<=59.9

Assessment Criteria

Course objectives assessed:	**Evaluation methods and criteria:**
Demonstrate knowledge of regular expressions (REGEX)	Students will create Python scripts (and run Python commands in the Shell mode) that use regular expressions to extract and/or modify information in the source file/string.
Demonstrate knowledge of web application program interfaces (APIs)	Students will create Python scripts (and run Python commands in the Shell mode) that use web APIs to upload web pages and/or resources locally.
Demonstrate the skills of processing information downloaded from the Internet	Students will create Python scripts (and run Python commands in the Shell mode) that read the source text of a downloaded web page and extract the desired information.
Demonstrate knowledge of web crawling	Students will create Python scripts that can access the information from pages and/or resources from web sites that do not have a custom API.

Course objectives assessed:	Evaluation methods and criteria:
Effective participation in class	Active participation is an essential part of the learning process and is required of all enrolled students. Participation means maintaining an active presence in the class by answering questions from the instructor, volunteering for and participating in class activities. Other elements of class participation include regular use of the target language in class with instructor and peers, constant demonstration of active listening when others speak, and collaboration with fellow students in in- class activities. Student participation will assist instructors in evaluating the language skills practiced in the course (listening /written comprehension tests, oral presentations, compositions, etc.). This grade will be impacted negatively if students fail to answer the instructor's questions or participate in class activities. Simply attending (or logging into) class does not constitute participation.
Work effectively in a team	The final project is a team project. Students will organize into teams and create a project that demonstrates their knowledge of the programming tools and techniques learned in class. The teams use the Internet and other resources to complete the project. There will be an oral presentation made to the class which will demonstrate students' learning experience from working in a group.

General Education Outcomes and Assessment

Learning outcomes:	Assessment method:
1. Demonstrate the ability to work collaboratively and independently on assignments in and outside of a classroom setting.	1. Classroom discussions, group assignments, and individual oral presentations.
2. Understand and employ both quantitative and qualitative analysis to solve problems.	2. Classroom discussion, group activities, group presentations, quizzes, tests, and final exam.

Learning outcomes:	Assessment method:
3. Develop reading, writing competencies, and listening skills.	3. Regular reading and writing assignments, individual and group presentation, classroom discussion. Each homework assignment requires writing.
4. Work with teams. Build consensus. Use creativity.	4. Group projects and presentations.

Academic Integrity

Students and all others who work with information, ideas, texts, images, music, inventions, and other intellectual property owe their audience and sources accuracy and honesty in using, crediting, and citing sources. As a community of intellectual and professional workers, the College recognizes its responsibility for providing instruction in information literacy and academic integrity, offering models of good practice, and responding vigilantly and appropriately to infractions of academic integrity. Accordingly, academic dishonesty is prohibited in The City University of New York and at New York City College of Technology and is punishable by penalties, including failing grades, suspension, and expulsion. The complete text of the College policy on Academic Integrity may be found in the catalog. The City College of Technology (CUNY CityTech) Student Handbook can be downloaded here: http://www.citytech.cuny.edu/current-student/docs/StudentHandbook.pdf

There is zero tolerance for cheating. The professor reserves the right to grade any quiz, exam, lab, exercise, assignment, or other academic work with a grade of zero if there is a reasonable expectation that the work is not an original work by the student.