NEW YORK CITY COLLEGE OF TECHNOLOGY/CUNY
Computer Systems Technology Department
Course Outline
CST 2312 – Information and Data Management I
(2 class hours; 2 lab hours; 3 credits)

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:
- Demonstrate knowledge of regular expressions
- Demonstrate knowledge of web APIs
- Demonstrate the skills of processing information downloaded from Internet
- Demonstrate knowledge of web crawling

Software: Python and Python Libraries

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.

Pre-requisites:
CST 1101 Problem Solving with Computer Programming

Required Text:
- Python for Everybody: Exploring Information by Charles R. Severance
  Free on-line version: http://www.pythonlearn.com/book.php#python-for-informatics
- Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython by W McKinney

Additional Reading Materials:
The instructor will identify several additional information resources during the semester including Internet resources, print material (handouts) and reference books.

Tools and On-Line Documentation:
- Python: https://www.python.org/
- Python Data Analysis Library: http://pandas.pydata.org/
- Python package for scientific computing: http://www.numpy.org/
- IPython command shell for interactive computing: https://ipython.org/
- Anaconda Data Science ecosystem: https://www.continuum.io/

Grading Procedure:

Midterm Exam	20%
Final Exam	30%
Lab Final Project	25%
Homework assignments	20%
In-class labs/quizzes	5%
	=====
TOTAL	100%

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

The grade distribution follows the information in the NYCCT Student Handbook (p.43).

Academic Integrity: During the course of the class you are required to follow the NYCCT Academic Integrity Standards described in the Student Handbook (pp.95 – 99)

NYCCT Student Handbook can be downloaded here: http://www.citytech.cuny.edu/current-student/docs/StudentHandbook.pdf.

Course Outline:

Week	Topics	Assignments and Readings
1	Review of Python Basics: variables, basic data types (String, int, float, Boolean, None), lists, selection, loops, functions	http://www.pythonlearn.com/html-270/book003.html
2	Introduction to IPython and IPython NoteBooks	Lecture notes Assignment 1
3	Introduction to Python tuples, and dictionaries Lab exercises: tuples, dictionaries	http://www.pythonlearn.com/html-270/book011.html
4,5	Introduction to regular expressions Lab exercises: regular expressions	http://www.pythonlearn.com/html-270/book012.html Assignment 2
6	Working with files; Using Libraries Lab exercises: importing libraries, using functions from imported libraries	http://www.pythonlearn.com/html-270/book008.html Assignment 3
7	**Midterm Exam:** Python tuples, dictionaries, regular expressions	http://www.pythonlearn.com/html-270/book010.html http://www.pythonlearn.com/html-270/book012.html
8	Introduction to Web APIs Web APIs using Python Lab exercises: using web APIs	Assignment 4
9	Regular Expressions and Crawling	
10	Reading CSV files, Selecting rows/columns	Assignment 5
11, 12	Pandas and NumPy: Introduction to data series, data frames Lab exercises: numpy arrays, pandas dataframes	
13, 14	Lab: Working on a project	
15	**Review, project presentation, final exam**	

ASSESSMENT CRITERIA: For successful completion of this course the student should be able to:

For the successful completion of this course a student should be able to:	Evaluation methods and criteria
Demonstrate knowledge of regular expressions	Students will create Python scripts (and run Python commands in the Shell mode) that use regular expressions to extracts and/or modify information in the source file / string.
Demonstrate knowledge of web APIs	Students will create Python scripts (and run Python commands in the Shell mode) that use web APIs to upload locally wen pages.
Demonstrate the skills of processing information downloaded from 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 form the web pages from the sites that do not have a custom web API.
Work effectively in a team	The final project is a team project. The students will group into teams and create a project that demonstrates their knowledge of the programming tools and techniques learned in class. Will use the Internet and other resources to complete the project. Also, there will be an oral presentation made to the class. It will include their learning experience in working in a group.

GENERAL EDUCATION LEARNING OUTCOMES/ASSESSMENT METHODS

LEARNING OUTCOMES	ASSESSMENT METHODS
1. Demonstrate the ability to work collaboratively and independently on assignments in and outside 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, final exam.
3. Develop reading, writing competencies, and listening skills.	**3.** Biweekly 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.

Bibliography:

- Charles R. Severance (2013). *Python for Informatics: Exploring Information*.
  Publisher: CreateSpace Independent Publishing Platform; 1 edition,
  ISBN-10: 1492339245
  Free on-line version: http://www.pythonlearn.com/book.php#python-for-informatics

- Wes McKinney (2017) *Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython.*
  Publisher: O'Reilly Media; 2 edition (October 20, 2017)
  ISBN-10: 1491957662

- Armando Fandango (2017). *Python Data Analysi*s
  Publisher: Packt Publishing - ebooks Account; 2nd Revised edition edition (March 27, 2017)
  ISBN-10: 1787127486

- Jake VaderPlas (2016). *Python Data Science Handbook: Essential Tools for Working with Data.*
  Publisher: O'Reilly Media; 1 edition (December 10, 2016)
  ISBN-10: 1491912057