NEW YORK CITY COLLEGE OF TECHNOLOGY/CUNY
Computer Systems Technology Department
Course Outline
CST 3512 – Information and Data Management II
(2 class hours; 2 lab hours; 3 credits)

Course Description:

This course continues to familiarize students with the informatics and intellectual tools necessary for students to become efficient and effective information users. The course covers topics related to the digital infrastructure, management and curation of data both structured (record-based) and unstructured (such as text). For structured data (data series, data frames), the course introduces Time Series data analysis and the basics of data visualization; for unstructured data (text), the course introduces text mining techniques. The course is project-based. During the course several case studies of using Information and Data Management are discussed. Based on these case studies the students develop their own projects where they choose a data set, use the appropriate data processing techniques for data analysis, and present their finding and conclusions reached using the information and data management and analysis techniques discussed in class.

Final Project:

The final project is a team project with an oral presentation. The teams choose an on-line data collection, obtain this data collection, analyze it, and create a visualization supporting interesting facts found in this data collection.

Course Objectives:

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

- Demonstrate knowledge of web APIs
- Demonstrate the skills of processing information downloaded from Internet
- Demonstrate knowledge of web crawling
- Demonstrate knowledge of basic text analysis techniques
- Demonstrate knowledge of basic visualization techniques

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 2312 Information and Data Management I

Required Text:

- Charles R. Severance. Python for Everybody: Exploring Information.
  ISBN-13: 978-1492339243
  Free on-line version: http://www.pythonlearn.com/book.php#python-for-informatics

- Wes McKinney. Python for Data Analysis: Data Wrangling with Pandas, NumPy, and IPython, 2nd edition
  ISBN-13: 978-1491957660

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%
Final Project	25%
Homework assignments	20%
In-class labs	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, 2	Review of Python Pandas basics: data series and data frames (CST 2302) using IPython Notes	
3, 4	More on data series and data frames	Assignment 1
5	Basics of visualization using MatPlotLib	Severance: Ch. 15
6,7	Introduction to text analysis	McKinney: Ch. 6, Ch. 7.3 Severance: Ch. 11
8	**Midterm Exam**; Introduction to Time Series data analysis, storing data over time	Assignment 2
9	Introduction to Time Series data analysis, storing data over time	McKinney: Ch. 11
10	Practice with a sample API (e.g., Citibike API), part I; Team project: proposals	https://www.citibikenyc.com/system-data
11	Practice with a sample API (e.g., Citibike API), part II; Team project: data collection from on-line resources	https://www.citibikenyc.com/system-data Assignment 3
12	Practice with a sample API (suggested in class); Team project: data collection from on-line resources	
13	Team project: data analysis and visualization	
14	Final exam review; **Final exam (the final exam is cumulative)**	
15	**Final Project Presentations / Discussion**	

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 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.
Demonstrate knowledge of basic text analysis techniques	Students will create Python scripts that analyze text information in the source files (e.g., find most frequent words, find most important entities mentioned in text, etc.)
Demonstrate knowledge of basic visualization techniques	Students will create Python scripts that visualize the input data.
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 Analysis*
  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