Syllabus for CMP 464/788: Topics in Computer Science: Data Science
Spring 2017

Instructor: Prof. Megan Owen
E-mail: megan.owen@lehman.cuny.edu
Office hours: Monday and Wednesday 9:30-10:50am, Gillet 137E or immediately after class
Course time: 1:50pm-3:30pm on Mondays (in Gillet 225) and Wednesdays (in Gillet 231)
Course website: http://comet.lehman.cuny.edu/owen/teaching/datasci/sp2017.html

Course Description: *4 hours, 4 credits*. Analyzing data sets to extract new insights. Topics: acquisition/scraping, data mining, storage, and visualization using Python. The emphasis of the course is on strengthening Python programming and analytic reasoning skills via analysis of real-world data. Prerequisites: Linear Algebra or Data Structures, proficiency in Python

Undergraduate vs. graduate course: The lectures will be the same, but the graduate coursework and exams will be at a higher level. Masters students must be enrolled in CMP 788 to receive graduate credit.

Textbook: Data Science from Scratch by Joel Grus. (ISBN 978-1491901427) Approximately $30. Available online from Amazon and O'Reilly

Grading: The grading for the course will be based on:

Problem Sets	30%
Quizzes	20%
Project	20%
Final exam	30%

You must take and pass the final exam to pass the course.

Homework: All problem sets should be submitted on Blackboard. **No late homework is accepted.** To receive full credit for a program, the program must perform correctly, must include comments, and be written in good style. Accompanying written analysis is expected to be written in standard written English (i.e. use complete sentences and proper grammar)

Quizzes: Instead of a mid-term examination, weekly quizzes taken outside of class will be used to assess mastery of the material.
• the quizzes have a 20 minute time-limit and are available on Blackboard
• later quizzes will include review questions as well as short programming exercises based on the homework
• a quiz may be taken multiple times before the deadline. **There are no make-up quizzes**.
• all quizzes are open book, open notes, and you may use the Python shell.

Project: A final project is required for this course. The grade for the project is a combination of the grades earned on the project milestones and the overall submitted project.

Final exam: The final exam is required and will be on Wednesday May 24 from 1:30-3:30pm. You must pass the final exam to pass the course.

Use of Technology & Blackboard: Coding for this course will be in Python, which is available from python.org. We will be using the numpy, matplotlib, and scipy libraries. These do not come with the default installations of python, but come with anaconda or visit the download pages of these libraries. If you would like to work through a browser, there are several web-based services. One that already has the libraries we are using in PythonAnywhere.

The at-home quizzes will be taken on the Blackboard system, provided by CUNY to all enrolled students. If you have not accessed Blackboard or are having difficulties, contact Blackboard Support in the Information Technology Division. You can also visit the Help Desk in the Computer Center (first floor, Carman Hall) in person. They can reset passwords and help with simple Blackboard issues.

Computer Access: Part of this course will use university computer laboratories. These machines are for work related to this course only and a code of conduct applies to computer use in the department and on-campus. Misusing university computers could result in losing your computer access for the rest of the term, making it exceedingly difficult to complete this course.

Honor Code: You are encouraged to work together on the overall design of the homework. However, for specific programs and homework assignments, all work must be your own. You should never copy code from the internet or another student. Both the student providing the code and the student copying in the code will receive 0 on the problem set or project. You are responsible for knowing and following Lehman's academic integrity code (available from the Undergraduate Bulletin, Graduate Bulletin, or the Office of Academic Standards and Evaluations). All incidents of cheating will be reported to the Vice President of Student Affairs.

Accommodating Disabilities: Lehman College is committed to providing access to all programs and curricula to all students. Students with disabilities who may need classroom accommodations are encouraged to register with the Office of Student Disability Services. For more info, please contact the Office of Student Disability Services, Shuster Hall, Room 238, phone number, 718-960-8441.

Learning Objectives:
At the end of the course, students should be able to:
1) Acquire data sets from multiple sources and write programs that can extract (scrape) the data into a usable form.
2) Use data mining to extract new insights about the data.
3) Understand basic storage techniques and constraints.
4) Analyze data using standard techniques from statistics and linear algebra.
5) Visualize data using popular Python modules.