CS 1656: Introduction to Data Science (CS 2056) – Fall 2016

Department of Computer Science, University of Pittsburgh

Lectures: (first one will be on Monday, August 29th)

Monday	1:00 pm – 2:15 pm	@	5129 Sennott Square
Wednesday	1:00 pm – 2:15 pm	@	5129 Sennott Square

Instructor: Prof. Alexandros Labrinidis

Office:	6105 Sennott Square	Office hours:	Monday:	2:15pm	–	3:30pm
Phone:	412-624-8843		Wednesday:	2:15pm	–	3:00pm
Web:	`http://labrinidis.cs.pitt.edu`					
Email:	`cs1656-staff@list.pitt.edu`					

Recitations: (first one will be on Friday, September 9th)

Friday 9:00 – 9:50 am / 1:00 – 1:50 pm @ 6110 Sennott Square

Graduate Teaching Assistant: TBD

Course Description: This course will provide an overview of data science technologies and techniques, offering a holistic view of the field, from data management & manipulation, to data analysis, and data presentation. The course will cover the main data management/querying paradigms (Relational/SQL, XML/XQuery, RDF/SPARQL, Graph/Cypher) along with information retrieval, data warehousing, data mining, data visualization, and other data analysis topics. The course will utilize Python as the default programming language and leverage existing libraries, as appropriate.

Prerequisites: A grade of C or better in CS 1501 is required (or permission of the instructor). Good working knowledge of Java and familiarity with Unix are assumed. Having passed a statistics course is highly encouraged.

Anti-requisites: Given the significant overlap with past offerings of CS1655, students who have already passed CS1655 will not be allowed to register for this class. The same applies for students who have passed CS1699 in the Spring 2015 term.

Class Web Page: `http://cs1656.org`

All handouts and class notes will be published on the class web page. You are expected to check this page frequently (at least twice a week).

Textbook: There is no single textbook with enough coverage of all the material that we will discuss in this class. We will rely on online references and also on O'Reilly's *Safari Bookshelf* for which the University has institutional access (i.e., you will not have to buy extra books).

Course Grading:

Assignments	48%	There will be 6 assignments (worth 6% - 10%), most of which will have a significant programming component (see important dates for deadlines).
Class participation	4%	For both lecture and recitations, including in-class quizzes. We will use the **Socrative** system to capture student responses and record attendance.
Midterm Exam	24%	Wednesday, October 19th, 1:00 pm – 2:15 pm (SENSQ 5129)
Final Exam	24%	Friday, December 16th, 2:00 pm – 3:50 pm (SENSQ 5129)

Important Dates:

- Thu, September 1, Assignment #0 released
- Thu, September 8, Assignment #0 **due**
- Thu, September 8, Assignment #1 released
- Thu, September 22, Assignment #1 **due**
- Thu, September 22, Assignment #2 released
- Thu, October 6, Assignment #2 **due**
- Thu, October 6, Assignment #3 released
- **Fall Break**: Class on Tue, October 18
- Wed, October 19, **Midterm Exam**
- Thu, October 27, Assignment #3 **due**
- Thu, October 27, Assignment #4 released
- Thu, November 10, Assignment #4 **due**
- Thu, November 10, Assignment #5 released
- **Thanksgiving Recess**: Nov 23 - Nov 27
- **Tue**, November 29, Assignment #5 **due**
- **Tue**, November 29, Assignment #6 released
- Thu, December 8, Assignment #6 **due**
- December 16, **Final Exam**

Class communications policies (NEW):

- **Mailing List** – All students will be automatically subscribed to the class mailing list, so that they receive time-sensitive announcements from the instructor and TA(s).
- **In-class student responses** – we will use the **Socrative** system (`http://www.socrative.com`) to capture student responses to questions and record attendance. It is crucial that you provide your Pitt user account name (e.g., xyz123) at the name prompt, to properly record your answers.
- **Email to instructor and TA** – **instead of email**, we will use the **Piazza** system (which is essentially a web-based bulletin board) for questions and clarifications to assignments. More instructions will be posted on the class web site.
- **Confidential Email** – in case you need to communicate with the instructor and TA outside of the Piazza system (i.e., for confidential matters), you should send email to `cs1656-staff@cs.pitt.edu`. We will make every effort to respond to all email requests within one business day at the latest. **Due to spam filtering, you should always use your pitt email address when sending email and include your full name**.

Cell Phone Use (NEW): Answering a cell phone or texting is very disruptive and hence any use of a cell phone to make or receive calls or text messages **is not permitted** in the class or recitation. Cell phones must be switched to silent mode and if you have a phone call which cannot wait until the end of the class, you need to step out of the class and then answer it.

Technology Policy (NEW): Since this is the 21^{st} century, the use of laptops, tablets, and other digital devices **is allowed** in class. **However**, when using digital devices in the classroom you must:

- **be mindful** – when you are emailing, tweeting, texting, surfing, etc, you are not paying attention. Research shows that no one can multitask that well, you included. Paying attention and taking good notes is essential to success in this course. Isn't that why you are here?
- **be respectful** – your use of digital devices should not distract other students in the class. It is unlikely that taking notes or searching class-related topics will be distracting to the other students. However, viewing videos of kittens or ice bucket challenges (gone well or gone wrong) will likely distract others. Complaints about inappropriate technology use in class will result in your privileges being curtailed or revoked.
- **be honest** – emailing, surfing, and the use of any other applications or technologies is not allowed during examinations. Doing so (unless explicitly allowed) is considered cheating in the exam and will be dealt accordingly.

Audio/Video Recording: To ensure the free and open discussion of ideas, students may not record classroom lectures, discussion and/or activities without the advance written permission of the instructor, and any such recording properly approved in advance can be used solely for the student's own private use.

Grading Policy: Unless explicitly noted otherwise, the work in this course is to be done independently. Discussions with other students on the assignments should be limited to understanding the statement of the problems (except when assignments are to be done in groups in which case it is expected of members of the same group to work together). **Cheating in any way, including giving your work to someone else, will result in an F for the course and a report to the appropriate University authority.** Submissions that are alike in a substantive way will be considered to be cheating by ALL involved parties. Please protect yourselves by only storing your files in private directories, and by retrieving all printouts promptly.

Students are expected to abide by the Dietrich School of Arts and Sciences' Academic Integrity code of conduct, posted at `http://www.as.pitt.edu/fac/policies/academic-integrity`

Grades can be appealed up to two weeks after they have been posted; no appeals will be considered after that time.

Assignment Policies (NEW - VERY IMPORTANT):

- All assignments must be submitted electronically.
- We will use `github` for assignment submissions – more instructions will be posted on the class web site. It is your responsibility to make sure your repositories are **private**. Doing otherwise will violate the academic integrity policy.
- It is crucial that you strictly adhere to the specifications for command-line arguments, input file format, and output file format, as specified in the assignment descriptions.
- It will be allowed (sometimes required) to use additional Python libraries in your assignments. However, you have up to five (4) calendar days to request if it would be ok to do so (using the piazza system). No additional libraries will be allowed to be included after that.

Academic Integrity Policy: Cheating/plagiarism will not be tolerated. Students suspected of violating the University of Pittsburgh Policy on Academic Integrity, noted below, will be required to participate in the outlined procedural process as initiated by the instructor. A minimum sanction of a zero score for the quiz, exam or paper will be imposed. (For the full Academic Integrity policy, go to `www.provost.pitt.edu/info/ai1.html`)

E-mail Communication Policy: Each student is issued a University e-mail address (username@pitt.edu) upon admittance. This e-mail address may be used by the University for official communication with students. Students are expected to read e-mail sent to this account on a regular basis. Failure to read and react to University communications in a timely manner does not absolve the student from knowing and complying with the content of the communications. The University provides an e-mail forwarding service that allows students to read their e-mail via other service providers (e.g., Hotmail, AOL, Yahoo). Students that choose to forward their e-mail from their pitt.edu address to another address do so at their own risk. If e-mail is lost as a result of forwarding, it does not absolve the student from responding to official communications sent to their University e-mail address. To forward e-mail sent to your University account, go to `http://accounts.pitt.edu`, log into your account, click on Edit Forwarding Addresses, and follow the instructions on the page. Be sure to log out of your account when you have finished. (For the full E-mail Communication Policy, go to `www.bc.pitt.edu/policies/policy/09/09-10-01.html`)

Late Policy: A late assignment will receive a deduction of 5 points if it is up to one day past the deadline and 15 points if it is up to two days past the deadline. Assignments that are past two days late will not be accepted.

Make-up Policy: Students are expected to be present for all exams and quizzes. Make-up exams will only be given in the event of an emergency, and only if the instructor is informed **in advance**. Failure to notify the instructor prior to missing an exam will result in a zero for the exam.

Final Exam Conflict Policy: In case you have a final exam conflict (i.e., have more than two exams scheduled on the same date during finals week), you need to notify the instructors of all classes involved in order to resolve the conflict by the sixth week of classes, according to the University policy (posted at `http://www.registrar.pitt.edu/classroomscheduling.html`).

Students with Disabilities:
If you have a disability for which you are or may be requesting an accommodation, you are encouraged to contact both your instructor and the Office of Disability Resources and Services, 140 William Pitt Union, 412-648-7890, as early as possible in the term. Disability Resources and Services will verify your disability and determine reasonable accommodations for this course. Their web site is `http://www.drs.pitt.edu`.

Religious Observances:
In order to accommodate the observance of religious holidays, students should inform the instructor (by email) of any such days that conflict with scheduled class activities **within the first two weeks of the term**.

Copyrighted Material All material provided through this web site is subject to copyright. This applies to class and recitation notes, slides, handouts, assignments, solutions, project descriptions, etc. You are allowed (and expected!) to use all the provided material for personal use. However, you are strictly prohibited from sharing the material with others in general and from posting the material on the Web or other file sharing venues in particular.

Outline: A detailed reading guide will be published on the web page, along with the class notes and additional online articles and resources. Time permitting, we will cover the following topics:

- Information Retrieval
- Data Mining
- Recommendation Systems
- PageRank / Network Analysis
- Data Warehousing
- Python Data Science Libraries
- SQL
- XML / XPath
- RDF / SPARQL
- Graph Databases / Cypher
- Data Visualization

[Last updated on August 29, 2016 at 11:44am EST]