CAS CS 105: Introduction to Databases and Data Mining

Boston University
Fall 2021

Syllabus

Description: Databases and other collections of data are everywhere. Retailers use data about customers and their purchases to make decisions that increase profits. Researchers analyze genomic data to find treatments for diseases. Policymakers analyze socioeconomic data to gain insights that guide their decisions. Online music and video services perform data mining to deliver customized recommendations. How does all this work? CS 105 examines how data is organized, processed and displayed. Topics include relational databases and the SQL query language, the writing of simple programs to process data, the principles of data visualization, and data-mining techniques for discovering patterns in data. At the end of the course, students apply the topics they have learned to a collection of data that interests them. Carries MCS divisional credit in CAS. This course fulfills a single unit in the following BU Hub area: Quantitative Reasoning II.

Prerequisites: none

Instructor
David G. Sullivan, Ph.D. (dgs@cs.bu.edu)
See the course website for instructor and TF office hours.

Teaching Fellow
Lina Qiu (qlina@bu.edu)

Meeting Times and Places
lectures: MWF, 1:25-2:15 pm, PHO 206
lab: a weekly, one-hour session; see your schedule for the time and location.

COVID-19-Related Class Expectations
To promote a safe learning environment, students must:

- comply with the University mandates for COVID-19 vaccination, screening, testing and contact tracing
- wear a face covering at all times during class.

Course Website: http://cs-people.bu.edu/dgs/courses/cs105
In addition, announcements and some course materials will be posted Blackboard.

Requirements and Grading

1. Nine problem sets (25% of the final grade)
2. A final project (10%). This will involve using the techniques covered in the course to organize and analyze a collection of data that interests you, to draw conclusions based on your analysis, and to present your results in a clear and compelling way. ***The project will be completed in teams of three.***
3. Three quizzes (25%)
4. Final exam (30%)
5. Preparation and participation (10%)

Collaboration Policy

You are strongly encouraged to collaborate with one another in studying the lecture materials and preparing for quizzes and exams.

You may discuss ideas and approaches to the assignments with others (provided that you acknowledge doing so in your solution), but such discussions should be kept at a high level, and should not involve actual details of the code or of other types of answers. **You must complete the actual solutions on your own.**

Academic Misconduct

We will assume that you understand BU's Academic Conduct Code:
http://www.bu.edu/academics/policies/academic-conduct-code

You should also carefully review the CS department's page on academic integrity:
http://www.bu.edu/cs/undergraduate/undergraduate-life/academic-integrity

Prohibited behaviors include:

- copying all or part of someone else's work, even if you subsequently modify it; this includes cases in which someone tells you what to write for your solution
- viewing all or part of someone else's work (with the exception of work that you and your partner do together on a pair-optional problem)
- showing all or part of your work to another student (with the exception of work that you and your partner do together on a pair-optional problem)
- consulting solutions from past semesters, or those found online or in books
- posting your work where others can view it (e.g., online)
- receiving assistance from others or collaborating with others during an exam, or consulting materials except those that are explicitly allowed.

Incidents of academic misconduct will be reported to the Academic Conduct Committee (ACC). The ACC may suspend/expel students found guilty of misconduct. ***At a minimum, students who engage in misconduct will have their final grade reduced by one letter grade (e.g., from a B to a C).***

Other Policies

Laptops: Students taking CS courses are expected to have a laptop capable of running a currently supported version of Microsoft Windows, Mac OS X, or Linux. See this page for more info: https://www.bu.edu/cs/undergraduate/undergraduate-life/laptops

Late problem sets: Problem sets must be submitted by the date and time listed on the assignment (typically by 11:59 p.m.). There will be a 10% deduction for submissions up to 24 hours late. **We will not accept any homework that is more than 24 hours late.** Plan your time carefully, and don't wait until the last minute so you will have ample time to ask questions and obtain assistance from the course staff.

Pre-lecture preparation: To help you prepare for lecture, you will typically be required to watch one or two short videos and to complete a short online quiz. Your work on these quizzes will not typically be graded for correctness, but it should demonstrate that you have adequately prepared for lecture. The pre-lecture quizzes must be submitted by the specified date and time; **late submissions will not be accepted**.

The *participation* portion of your grade will be based on your completion of the pre-lecture quizzes and in-lecture questions, and on your consistent participation in the lab sessions. You will receive full credit for participation if you answer at least 85% of the online lecture questions and if you participate in at least 85% of the lab sessions. If you complete x% of the questions or participate in x% of the lab sessions for a value of x that is less than 85, you will get x/85 of the possible points.

The final exam will replace your lowest problem-set grade if doing so helps your final grade. The final exam will also replace your lowest quiz grade if doing so helps your final grade. The final grades are *not* curved. The performance of the class as a whole is taken into account when assigning letter grades, but this can only improve your grade, not harm it.

Extensions and makeup exams will only be given in *documented* cases of serious illness or other emergencies.

You cannot redo or complete extra work to improve your grade.

Incompletes will not be given except in extraordinary circumstances.

Course Materials

- **Required:** CS 105 Coursepack. This contains all of the lecture notes for the course. More detail will be provided in class and in Lab 0.
- **Required:** We will be using the Top Hat Pro platform. More detail will be provided in class.

Schedule (tentative)

week	lecture dates	topics, exams, and special dates
0	9/3	Course overview and introduction *No labs this week*
1	9/8, 9/10	Database fundamentals The relational model: How is data organized in a typical database?
2	9/13, 9/15, 9/17	SQL: a language for creating, modifying, and extracting data from a database *9/17: last day to add a class* **Problem Set 1 due on 9/16**
3	9/20, 9/22, 9/24	SQL (cont.) **Problem Set 2 due on 9/23**
4	9/27, 9/29, 10/1	SQL practice Programming in Python: fundamentals, loops **Problem Set 3 due on 9/30**
5	10/4, 10/6, 10/8	Python: working with numbers, making decisions **Quiz 1 on 10/6** *10/7: last day to drop without a 'W'* **Problem Set 4 due on 10/8 (Fri)**
6	**10/12**, 10/13, 10/15	Python: working with strings/text and lists; accessing a database *No lecture on 10/11 (Indigenous Peoples' Day)* *Lecture on 10/12 (Mon. schedule)* **Problem Set 5 due on 10/14**
7	10/18, 10/20, 10/22	Python: working with data stored in a text file **Problem Set 6 due on 10/21**
8	10/25, 10/27, 10/29	Data visualization: how to create compelling and useful data graphics **Problem Set 7 due on 10/28**
9	11/1, 11/3, 11/5	Data mining: what is it? How can a computer learn? *11/8: last day to drop a class with a 'W';* *last day to designate a course Pass/Fail* **Quiz 2 on 11/3**
10	11/8, 11/10, 11/12	Data mining: learning to categorize things **Problem Set 8 due on 11/11** **Final-project proposal due on 11/14**
11	11/15, 11/17, 11/19	Data mining: finding patterns in numeric data; discovering arbitrary relationships in data
12	11/22	Data mining: preparing data for mining **Problem Set 9 due on 11/22** *No lectures on 11/24, 11/26 (Thanksgiving)* *No labs this week*

13	11/29, 12/1, 12/3	Case studies and/or project sessions **Quiz 3 on 12/1** **Work on final project**
14	12/6, 12/8, 12/10	Case studies and/or project sessions **Final projects due on 12/10** *12/11-13: Study period*
15		**Final exam: Sat, 12/18, 12:00-2:00 pm** *The exam must be taken at this time, so please plan your travel accordingly!*