DATA/COMP 200: Database Systems and Data Management for Data Analytics

Fall 2022

(Last updated August 31, 2022; syllabus is subject to change)

Instructor: Professor Eren Bilen
Office: Rector North 1309
Email: bilene@dickinson.edu
Phone: 717-254-8162
Office Hours: Monday 4:30-5:30pm
             Tuesday 3:00-4:00pm
             Wednesday 9:00-10:00am
             or by appointment
QRA: Zimeng Liu, liuz@dickinson.edu
Office Hours: Wednesday 1:45-2:45pm
             Thursday 1:45-2:45pm
Location: Rector North 1311

Class: Tome 121
Monday and Thursday
3:00-4:15pm

Class Notes and Other Required Materials

- DATA 200 Course Pack by Dick Forrester. A pdf version will be provided.
- Introduction to Data Systems: Building from Python (1st edition), by Bressoud and White, Springer, 2020. Available at bookstores, or online on Amazon.
- (Optional) Ace the Data Science Interview by Huo and Singh
- DataCamp: This course will make extensive use of DataCamp.com, which is an online learning platform specifically designed for data science. You will be enrolled in our course group on the DataCamp website and should receive an e-mail to your Dickinson account with a link to register on their site. Please note:

  - You will be provided with six free months of access to all the content on Data Camp. There is an incredible selection of content available, and you are encouraged to take advantage of this free access.

  - You are not obligated to purchase any services from DataCamp. At the end of the six months, you will need to purchase a subscription if you wish to continue to use the site, unless it is made available to you through another data analytics course at Dickinson.

- Access to a computer to install and use Anaconda.
- Course webpage: Github

Course Overview

A comprehensive introduction to the management and manipulation of database systems as it applies to data analytics. Topics related to data query languages to relational

databases and NoSQL data systems will be covered, as well as the access and acquisition of other structured and unstructured data repositories available across the Internet. An understanding of techniques for transforming and restructuring data representations to allow for analysis will also be addressed. At the end of the course, students will be able to

- understand the tabular data model, the relational data model, and hierarchical data model
- retrieve data using Structured Query Language (SQL)
- understand the client-server model for communication
- acquire data from a spectrum of external systems, ranging from structured to unstructured systems using APIs
- manipulate unstructured data into meaningful representations
- utilize the Python language as it applies to data analysis

We will make extensive use of Python and Anaconda distribution to generate graphical and numerical representations of data and complete data obtaining and processing techniques via SQL before analyzing data. We will revisit a number of concepts that you have learned in Data 180, and see their implementation in Python. Python is an extremely popular language among data scientists, and together with R, it gives a huge amount of flexibility as a computational tool in addition to being an immediate resume builder!

Course Policies

Attendance Policy: This course will be taught in person in Tome 121. Students are expected to attend all in-class meetings, which occur on Tuesdays and Thursdays from 3-4:15pm EDT. While I will not take formal attendance, it is important for you to attend the class meetings and take notes. If you will be unable to attend a class meeting for any health-related issues or other emergencies, please contact me beforehand so that arrangements can be made.

Use of Laptops, Tablets, and Phones: Laptops and tablets are permitted for note-taking during this course. In exchange for trusting you to use these devices, I ask that you not use them as distractions. I maintain the right to change this policy for individual students or for everyone if these tools become a problem during class. Phones are not permitted and should be put away in silent mode.

Grading: Your course grade is based on two closed-book midterms, a take-home final exam, and homework assignments.

Midterm 1 (20%):	October 13
Midterm 2 (20%):	November 17
Take-home Final (20%):	by December 16, 11:59am
Homework (40%):	Due dates TBA

While I will not be giving extra credit in this course, I will drop your lowest homework. I expect there to be 12-13 total assignments (depending on course pacing). Occasionally, an assignment may be weighted to count as two assignments (because of the complexity or length), this will be clearly indicated when it is assigned.

The following scale will be used to determine your final grade:

Score	Letter	GPA	Score	Letter	GPA
$93 \geq x$	A	4.0	$73 \leq x < 77$	C	2.0
$90 \leq x < 93$	A-	3.7	$70 \leq x < 73$	C-	1.7
$87 \leq x < 90$	A-	3.3	$67 \leq x < 70$	D+	1.3
$83 \leq x < 87$	B+	3.0	$63 \leq x < 67$	D	1.0
$80 \leq x < 83$	B	2.7	$60 \leq x < 63$	D-	0.7
$77 \leq x < 80$	C+	2.3	$x < 60$	F	0.0

Make-up Exams: There will be no make-up exams unless a student must be away from campus on university business or due to an emergency. The student must provide documentation. If an emergency arises, you must inform me as soon as possible. Once you provide me an official documentation related to the emergency/university business, you may schedule a make-up exam. Warning! It is absolutely essential to provide me documentation. You will receive 0 if you are unable to get an official documentation. Therefore, you should definitely not skip a test if your situation cannot produce documentation.

Homework: Homework assignments will be posted on course Github page as a Jupyter Notebook template file on which you will insert your solutions. Due dates will be provided for each assignment. You will turn in your assignments as a Jupyter Notebook file and as a Markdown file via a pull request from your private GitHub.com repository which is a clone of the class master repository. (You will need to set up a GitHub account if you do not already have one.) You will be sent an invitation link for each assignment. After accepting the assignment, your private repo where you push your files in will automatically be created. Prior to pushing your submission files to your repository, make sure to hit `Kernel > Restart & Run All` on your Jupyter Notebook, and save your notebook with `File > Download as > Notebook (.ipynb)` and `File > Download as > Markdown (.md)` Make sure to unzip all output, figures, images before pushing your commit. You will receive a 20% penalty if your code cannot get executed because of errors, so make sure to hit `Run All` and verify your code runs with no issues. Email submissions will not be accepted. Late assignments will not be graded.

You are encouraged to work in teams, but your submissions must be individual. It is important that you must understand and be able to explain every part of the code you are submitting. I do not want to see a bunch of copies of identical code. I do want to see each of you learning how to code these problems so that you could do it on your own. Homework assignments will require the use of Python and the Anaconda Package; you will want to obtain access to a computer with Anaconda installed during the first week of classes; Anaconda package is installed in Tome 121 and various labs in Tome Hall.

Take-home Final: The course will include a final data project in lieu of a final exam that will be due Friday, December 16 at 11:59 am EDT. More information will be posted later in the semester.

Getting Help

Office Hours: I will be holding three hours of office hours each week. Please see the first of page of the syllabus for my hours. I am also available by appointment. If there is a conflict and you are unable to make it to any of my hours, please feel free to send me an email. My availability outside office hours is not guaranteed, however I devote my attention fully to you during my office hours. Therefore, I highly encourage you to come to my office hours and ask questions.

Quantitative Reasoning Associate: This semester, we are fortunate to have a Quantitative Reasoning Associate (QRA) working with us. A QRA is a fellow student who completed this course in the past and will be helping us as a course facilitator and student mentor. This semester, the QRA for our course is Zimeng Liu. She will be holding office hours during the hours on the first page of the syllabus. Location is TBA.

In addition, Zimeng will host a study session before each exam, which will be announced closer to exams.

Quantitative Reasoning Center

Dickinson College provides additional support for students taking courses with quantitative content across the curriculum through the Quantitative Reasoning (QR) Center. For the Fall 2022 semester, the QR Center will offer tutoring for DATA 200, in addition to general quantitative support. You are strongly encouraged to make an appointment with them. `Click here` to access the QR Center webpage.

Please visit `dickinson.mywconline.com` to make an appointment. Then, access the drop-down menu under "limit to" at the top of the scheduler and select DATA 200. This will restrict the tutor list and schedule to only those tutors approved for this course. When you make your appointment, please also paste or upload your assignment and any work that you have done.

Other Important Information

Referencing the Work of Others: When submitting your work, you must follow common-sense ground rules. External sources may only be used to improve your own understanding of the material. When you write your solutions, you should do it on your own without the direct help of any external sources, and certainly should not write down anything that you do not understand. If you do use external references, please be sure to cite them. Failure to cite references will be treated as academic dishonesty.

Respect for Intellectual Property: It is important that you be aware of and respect the intellectual property rights of others. Unless explicitly stated otherwise, all materials available on the Internet, in libraries, and elsewhere are considered intellectual property and can only be used with the permission of the owner. Specifically, with regards to this class, you should not share any of the course materials, including homework answer keys, with others, even after the completion of the course.

Statement on Disabilities: Dickinson values diverse types of learners and is committed to ensuring that each student is afforded equitable access to participate in all learning

experiences. If you have (or think you may have) a learning difference or a disability – including a mental health, medical, or physical impairment that would hinder your access to learning or demonstrating knowledge in this class, please contact Access and Disability Services (ADS). They will confidentially explain the accommodation request process and the type of documentation that Dean and Director Marni Jones will need to determine your eligibility for reasonable accommodations. To learn more about available supports, go to www.dickinson.edu/ADS, email access@dickinson.edu, call (717) 245-1734, or go to the ADS office in Room 005 of Old West, Lower Level (aka "the OWLL").

If you have already been granted accommodations at Dickinson, please follow the guidance at www.dickinson.edu/AccessPlan for disclosing the accommodations for which you are eligible and scheduling a meeting with me as soon as possible so that we can discuss your accommodations and finalize your Access Plan. If test proctoring will be needed from ADS, remember that we will need to complete your Access Plan in time to give them at least one weeks advance notice.

SOAR: Academic Success Support: Students can find a wealth of strategic guidance by going to www.dickinson.edu/SOAR. This website for SOAR (Strategies, Organization, and Achievement Resources) includes apps, tips, and other resources related to time management, study skills, memory strategies, note-taking, test-taking, and more. You will also find information aimed to helps students "SOAR Through Academic Challenges," as well as a schedule of academic success workshops offered through Academic Advising. If you would like to request one-on-one assistance with developing a strategy for a manageable and academically successful semester, email SOAR@dickinson.edu.

Course Outline: Below is a list of topics to be covered in this course. There may be adjustments on the list during the semester depending on progress. Any adjustments will be announced and this syllabus will be updated.

- Topic 1: Introduction to Database Systems & Jupyter

- Topic 2: Introduction to Numpy

- Topic 3: File systems and paths

- Topic 4: Introduction to Pandas

- Topic 5: Visualization with Matplotlib

- Topic 6: Tabular Model

- Topic 7: Introduction to Relational Models

- Topic 8: Coding with Relational Models: SQL

- Topic 9: Table joins

- Topic 10: Introduction to APIs and Web-scraping

Important Dates for the Fall 2022 Semester

Last Day to Add/Drop or Change to/from Pass/Fail	Friday, September 2
Mid-Term Pause	5 pm, Friday, October 14 thru 8 AM, Wednesday, October 19
Course Request Period for Spring 2022 Semester	Monday, October 31 thru Wednesday, November 2
Thanksgiving Vacation	5PM, Tuesday, November 22 thru 8 AM, Monday, November 28
Last Day to Withdraw from a Course with a "W" grade	Tuesday, November 22
Classes End	Friday, December 9
Reading Period Days	December 10, 11