DATA 1050 SYLLABUS
FALL 2021
BROWN UNIVERSITY
SAMUEL S. WATSON

EMAIL	sswatson@brown.edu
TAS	Meera Kurup (HTA), Lukas Kania, Bangxi Xiao, Yue Wang, and Ian Acosta
WEBSITE	https://data1050.github.io
CLASS MEETING	MWF from 13:00 to 13:50 in Friedman 201
OFFICE HOURS	Mondays and Fridays, 14:00 to 15:00
COURSE DESCRIPTION	We will delve into a range of computer science topics which are relevant for data scientists. Students will be able to...

1. Import and manipulate data in a variety of formats
2. Discuss how data is managed within organizations
3. Describe how computers work at a basic level and reason about the implications of these hardware details for how we build software
4. Take advantage of productivity-enhancing features of development environments (VS Code and Jupyter)
5. Perform basic operations using the command line (Bash)
6. Version control their software (Git)
7. Solve programming exercises (Python)
8. Create data visualizations using dashboarding software (Superset)
9. Describe the relational data model and devise SQL schema appropriate to a given use case
10. Set up a SQL database and write SQL queries to perform basic data manipulation tasks (PostgresQL with Supabase)
11. Discuss the advantages and disadvantages of noSQL databases, set up and use a noSQL database (MongoDB).
12. Solve exercises on data structures and algorithms (including abstract data types, asymptotic notation, sorting and binary search, graph algorithms, and database algorithms).
13. Describe the paradigmatic use cases for graph databases (neo4j) and streaming databases (Kafka), and perform basic tasks using those databases.
14. Build systems which can perform computations in parallel across multiple nodes (PySpark)
15. Get data from the web via scraping or interacting with REST APIs.

16. Deploy a dashboard-style website which draws from some data source and updates live.

While this list of learning goals is sweeping and ambitious, we are not prioritizing the development of highly advanced skills in any one of these areas. Instead, we are looking to give you a solid grounding in the fundamentals, as well as practical experience with getting started and knowing where to look to develop your skills further when the opportunity arises. We will also be leveraging opportunities to minimize hassle on your part, so that you can focus on high-traction learning activities.

STUDENT SATISFACTION & INCLUSION

Our top priority is for you to have an excellent learning experience in this course. We intend to set clear learning objectives and equip you with the right tools to achieve them. We invite comments, criticisms, concerns, and suggestions at any time. If you perceive that you are not doing as well as you'd like, please see us right away. We can help with data engineering ideas, of course, but we are also happy to help you troubleshoot your approach to studying, problem solving, etc. We will grant you the assumption that you are doing your best to learn, and we have zero interest in making you feel judged for where you are in the learning process. *You should have every expectation that you can translate sustained hard work into a high level of course success*.

Data science is a new, multidisciplinary field which benefits from the perspectives and contributions of those from diverse disciplinary and personal backgrounds. Your instructional team is committed to fostering an inclusive and intellectual environment that supports a diversity of thoughts, perspectives and experiences, and honors your identities (including nationality, gender, race, class, sexuality, religion, etc.).

PRISMIA

Prismia (https://prismia.chat) is a website built by this course instructor for purposes of facilitating and enhancing this course and others. As an effort to reduce the number of websites students have to manage for this course, we will be doing all major course operations (class, assignments, and message board) in Prismia.

The course join link is available on the internal site linked from data1050.github.io.

CLASS SESSIONS

Class will be mediated by Prismia and Zoom (for remote students). Class content will present as a chat thread which includes figures, executable code, multiple-choice questions, and open response questions. You should expect class to be very interactive and learning-intensive.

Prismia data on student participation will be used to generate a participation score. Students who answer all or almost all of the questions asked on Prismia will receive full participation credit. You can miss up to 3 days with no penalty. Others will receive points in proportion the proportion of question answered.

Students who are not able to participate in live class sessions should request accommodation.

DRILLS We will provide simple exercises to help you practice skills developed in class. These will be designed to be useful for preparing for homework, exams, and job interviews. These are provided through a dedicated Prismia page which will allow you to submit new questions that will be vetted and potentially edited by the instructor to make available for other students to use.

ASSIGNMENTS Written assignments will be due at 6 PM on each Monday, with a 2-hour grace period. You can grant yourself an extension of up to 3 days (directly on the assignment in Prismia, under the Settings icon), but your total number of extension days

HOMEWORK POLICY There are no dropped homework assignments. Dates are coordinated with DATA 1010 and 1030, so curricular conflicts should not be an issue. In event of medical or family emergency, contact us for accommodation.

EXAMS We will have two exams (a midterm and a final) in which students will be asked to perform tasks that have been practiced in class and on homework. These exams will be given over a 3-hour block. You will have an opportunity on the final to improve your grade on topics on which you missed points during the midterm.

GRADING Your final grade will be determined as follows.

Participation	20%
Exams	20%
Homework	40%
Projects	20%

PREREQUISITES Students should have a reasonable degree of facility with some programming language coming into the course, and students should have at least some familiarity with Python.

COLLABORATION AND ACADEMIC INTEGRITY You may collaborate on solving homework problems, but you must write your solutions entirely by yourself, and you may not do so with reference to notes taken while working in a group. Writing solutions based on notes which represent the ideas of others short-circuits the exercise and impedes your learning. On each homework submission, write a statement on the front page either listing collaborators or confirming that you did not collaborate. Using the internet to look up solutions to homework problems is not acceptable, although of course you may freely use any available resources to learn the material more generally. You may not copy-paste code from any source without attribution.

Please view our Collaboration Policy Agreement form and fill out the Google Form. Both the agreement and form can be accessed on the course website and Prismia.

MESSAGE BOARD You can ask questions in office hours or on the Prismia Message Board for this course. Start a new thread (which may be private or public) and type in a message to ask a question or make a comment. Generally speaking, a thread should be

public unless it includes spoilers (like your code). You have the option to post anonymously.

COURSE-RELATED WORK EXPECTATIONS

Students will meet 3 hours per week in class (43 hours total). Weekly written assignments will take about 7 hours per week (97 hours total). In addition, the projects are expected to require about 20 hours each (40 hours).

DISABILITY SUPPORT

Please inform us if you have a disability or other condition that might require modification of these procedures. We are happy to accommodate your learning needs. You should also contact the Student and Employee Accessibility Services at 401-863-9588 or SEAS@brown.edu.