Syllabus for CDS DS 121: Foundations of Data Science II in Fall 2023

Instructor: Mayank Varia (varia@bu.edu)
TAs: Abhishek Tiwari and Dhun Jayswal
In-class midterm exams: October 5 and November 9

Lecture: Tues Thurs 1:30-2:45pm in WED room 130
Discussion: Mon 9:05-9:55 / 12:20-1:10 / 1:25-2:15 / 2:30-3:20
Final exam: Yes, following BU's final exam schedule

This syllabus contains information you need to know about DS 121. Please **read this document carefully** in the first week of class, familiarize yourself with how the course works, and maintain that familiarity throughout the semester. You are responsible for adhering to the course policies at all times, especially the academic code of conduct, plagiarism, AI, and collaboration policies.

Course Information

Course Description: DS 121 is the second in the three-course sequence (DS 120, 121, 122) that introduces students to theoretical foundations of Data Science. DS 121 covers an introduction to key concepts from Linear Algebra (vector space, independence, orthogonality and matrix factorizations). The DS theme running through the course is exploratory data analysis, enabling a better understanding of the data at hand. The course will link mathematical concepts with computational thinking, specifically through the use of problem sets that require students to answer mathematically-posed questions using computation. This course fulfills a single unit in each of the following BU Hub areas: Quantitative Reasoning I, Digital/Multimedia Expression, Critical Thinking.

Prerequisites: DS 120 (or equivalent) is a prerequisite, and DS 110 (or equivalent) is a corequisite.

Course websites: We will use two websites in this course, Piazza and Gradescope. *Please sign up for both immediately!*

- *Piazza* (https://piazza.com/bu/fall2023/ds121) contains the weekly course schedule, homework assignments and required reading; you should review the course schedule at least once per week. Also, course announcements will be made using Piazza; you are responsible for knowing these announcements, so please register for Piazza using an email address that you check regularly. We also encourage you to use Piazza as a resource to ask scientific questions to the instructors and your peers, and recommend that you mark posts as public unless they contain a personal question or reveal a partial solution to an assignment.
- *Gradescope* (https://www.gradescope.com/courses/605010, entry code 6GZXBV) is the website that you must use to submit completed homework assignments and to respond to in-class questions.

Course textbooks: We will use (portions of) 5 textbooks during this class, all of which are available to download for free using the links in the Piazza resources tab: https://piazza.com/bu/fall2023/ds121/resources. (BU login may be required.)

Course meetings: The course instructor is Prof. Mayank Varia. I will lead lectures on Tuesdays and Thursdays at 1:30-2:45pm in WED room 130. All lectures will be livestreamed over Zoom; if you cannot come to class in person (e.g., due to illness), you can join the Zoom livestream or review the lecture notes and video on Piazza after class.

Discussion labs are held Mondays at 9:05-9:55am (in CDS 164), 12:20-1:10pm (in BRB 121), 1:25-2:15pm (in CDS 264), and 2:30-3:20pm (in CGS 311). We actively encourage questions and interaction during all lectures and discussion labs.

Office hours: Students are welcome and encouraged to visit office hours. We will post the date, time, and location of all office hours on Piazza, and will also announce in class and on Piazza if the office hour times ever change. If you want to meet with me (the professor) but cannot make the scheduled times, then send me a private Piazza note with at least 3 suggestions for times that you are available to meet and we will find an alternate time.

Course Assessment

Homework: We will assign homework weekly, except on the weeks of midterm exams. Homework will typically be assigned on a Thursday and due on the following Friday at 6pm, unless otherwise specified. Homework assignments will include two types of questions: pen-and-paper mathematical questions that require you to calculate an answer or explain a concept, and computer-based data science questions that require you to write computer code in Python to analyze a dataset. I recommend using Google Colab (https://colab.research.google.com) for the programming questions, although it is not required.

Homework assignments are assessed on the *correctness* and *clarity* of your solutions; the graders are looking for positive evidence that you understand the relevant concepts. Before submitting an assignment, you should crop down to the pertinent work, document all code, describe your thought process, display any relevant outputs, clearly mark your final answer, and explain your findings. For instructions on how to submit an assignment, see: https://help.gradescope.com/article/ccbpppziu9. Always review your file after uploading to check that it is legible and complete; we can only grade what we can see.

Exams: There will be *three exams* in this course: two in-class midterms on October 5 and November 9, and a final exam held on the university's Final Exam Schedule (tentatively, December 19 at 3-5pm). *Reserve these dates on your calendar now!* All exams are cumulative: they may cover any material from lectures, discussion lab sections, homework, and required textbook reading.

Participation: To earn full participation credit, you must attend at least 11 out of the 14 discussion labs.

Overall course grade: Your grade in this course is based on your grades on the homework assignments (40% total, with all weighted equally after the lowest one is dropped), the two midterms (15% each), the final exam (20%), and participation (10%). I will determine grade cutoffs after all assignments and exams have been graded. Grade cutoffs will take into account my assessment of the difficulty level of the assignments and exams, and my assessment of what is expected for each letter grade.

Course Policies

Academic code of conduct: You must read and adhere to BU's Academic Code of Conduct, which is available here: https://www.bu.edu/academics/policies/academic-conduct-code/. Please familiarize yourself with this code, its definitions of misconduct and plagiarism, and its sanctions. Violations of this code will result in *receiving a score of 0 on the homework or exam*, and may be grounds for referral to BU's Academic Conduct Committee. If you have any questions about the policy, you must ask me in person or via private Piazza note *before* taking an action that might be a violation.

Plagiarism: All written work in this course must be original to you. If you consult outside texts, or other forms of assistance, cite these sources in the proper format—at a minimum, include the author, title, and website link for all external sources (books, journals, lectures, web sites, AI). We are required to report all suspected cases of plagiarism to the Academic Dean for review.

Academic integrity in computing coursework has some special aspects. Please review the examples of plagiarism as provided by the BU CS department: https://www.bu.edu/cs/undergraduate/undergraduate-life/academic-integrity/.

Generative AI policy: All submitted work in this course must conform to the CDS Generative AI Assistance Policy, which you can read at https://www.bu.edu/cds-faculty/culture-community/gaia-policy/. Also, keep in mind that AI tools are often wrong!

Collaboration policy: The goal of homework and project assignments is to learn. Hence, I encourage you to use any and all resources that can help you to learn the material: computers/calculators, Piazza, lecture notes, textbooks, other websites, and your fellow classmates. That said, please always obey the following rules:

- You *cannot* copy solutions from anyone else, or give your solutions to a classmate to copy.
- You also *cannot* actively search for the solutions to the homework questions on the Internet or in any other source.
- Your submission *must* list (a) names of all classmates you worked with, (b) all websites you used besides the ones listed in the lecture notes or textbooks, and (c) all code that you used from other sources, including the exact prompts and responses from any AI tool. Taking ideas without attribution will be considered plagiarism.

By contrast, the goal of the exams is for you to show me what you have learned, so any form of collaboration is strictly prohibited. Computers and notes are also forbidden during exams unless I explicitly state otherwise. (That said, I encourage you to collaborate with classmates when studying lecture materials and preparing for the exams.)

Accommodations: BU strives to be accessible, inclusive, and diverse in our facilities, programming, and academic offerings. Your experience in this course is important to the teaching staff. If you have a disability or believe that you require a reasonable accommodation, please meet with BU Disability and Access Services as soon as possible at the beginning of the semester. Their office is at 25 Buick Street, Suite 300, and they can be contacted at 617-353-3658. Requests for accommodations are sent by that office to the Academic Dean who approves and returns them. Disability and Access Services then forwards them to the instructor.

Learning environment: This course seeks to be inclusive of people of all genders, races, cultures, abilities, and sexual orientations. Please respect your fellow classmates and contribute toward a positive learning environment for everyone. While I actively encourage discussion and debate on ideas, I won't tolerate criticism of other people. Also, while you can use a computer for note-taking, do not use your laptop or cellular phone in class for web surfing, sending messages, or anything else that can cause a distraction to your fellow classmates.

Absence policy: This course follows BU's policy on absences for religious observance. Otherwise, students should attend the lectures and discussion lab sections, either in person or virtually via Zoom. Due to ongoing public health concerns: if you feel sick, please err on the side of caution and attend via Zoom or review the lecture notes and video afterward on Piazza.

Late work policy: You are responsible for submitting homework electronically on Gradescope by the stated due date and time. Assignments will be accepted on Gradescope up to 6 hours late for a 20% grade reduction. Later assignments will not be accepted, except in cases of long-term emergencies (e.g., family, medical); if this applies to you, inform us as soon as you are able. To accommodate the possibility of missed assignments, I will drop the lowest homework grade.

Make-up exam policy: If you have a valid conflict with a midterm exam, you must send me a written note *as soon as you are aware*, and with a *minimum of 2 weeks notice* (barring extenuating circumstances) so we can arrange a make-up exam. The final exam can only be rescheduled in accordance with the university policy: https://www.bu.edu/reg/calendars/final-exams/policy/.

Regrade policy: You have the right to request a regrade of any homework or exam question. All regrade requests must be submitted via Gradescope, and must describe a factual error in our assessment. If you request a regrade for one question, then we have the right to review the entire homework or exam. Beware that this may potentially result in a lower grade.