INFO 3401: Information Exploration

M/W/F, 2:00-2:50, CASE E250

Prof. Danielle Albers Szafir
danielle.szafir@colorado.edu

Office Hours: TBD & by appointment, TLC 276

Course Description

Information empowers people to build deeper understandings of the world and make more informed decisions. However, the increasing volume and variety of available information makes it hard for people to make sense of that data. This course will allow you to build the skills necessary to work with stakeholders to explore and build novel insights through data. You will gain hands-on experience with different tools and techniques for exploring information, including statistical methods, qualitative analyses, and visual analytics. You will learn how to generate and synthesize new findings from data, combine information from multiple sources, and identify questions and findings that are directly relevant to people.

Course Structure

The intention of this class is to help you combine the skills you've built in the Computational Thinking, Quantitative Thinking, and Foundations courses. By the end of this class, you will have gained experience applying the methods you've used to work with small datasets to larger and more complex problems. As such, the course will rely heavily on a combination of video lectures to be watched before class, in-class lectures and discussion, and in-class hands-on activities where we will work through data problems using the concepts from lecture.

While this course is not a programming course, many in-class activities and projects will require programming in-class. As a result, please plan on regularly bringing a laptop to class that you can use to develop and run your own programs. We will primarily use Python; however, time permitting, we will also likely work with other languages such as R. If you do not have access to a machine that will support your work in this class, please let Dr. Szafir know by the end of Week 1.

Due to the use of video lectures, there will be no required textbook for this class. Additional readings may be provided to support individual concepts, but will be made available through Canvas. The discussion board on Canvas will be actively monitored to support questions that arise from video lectures and we will additionally take time during each class period for discussions and clarifications. If you find yourself in a situation where you would like additional readings to support the lectures, please email Dr. Szafir for additional materials. Similarly, if you find resources online that are particularly useful, please feel free to share these with the class on the Canvas discussion board.

Topic Schedule: (Subject to Change)

August 27-31	• Welcome & Course Overview • Tools of the Trade
September 3-7	• Tools of the Trade, Continued • History of Data Science • The Data Science Pipeline
September 10-14	• Basics of Project Management • Scoping & Requirements • Data Collection Methods
September 17-21	• Data Collection Methods, Continued • Polling, Samping, and Approximation • Data Types
September 24-26	• Common File Types • Working with Databases
October 1-5	• Data Wrangling Strategies • Exam #1
October 8-12	• Descriptive Statistics & Distributions • Dealing with Noise & Outliers • Inferential Statistics
October 15-19	• Inferential Statistics, Continued • Visual Analytics
October 22-26	• Probability • Linear & Logistic Regression • Multiple Regression
October 29-November 2	• Multiple Regression & Prediction, Continued • Exam #2
November 5-9	• Naive Bayes • Clustering Algorithms
November 12-16	• Evaluation Metrics • Bias-Variance Trade-Off • Distance Metrics
November 19-23	• Decision Trees • Feature Selection • Ensemble Methods
November 26-30	• Fall Break, No classes
December 3-7	• Text and/or Network Analysis Fundamentals
December 10-14	• Final Project Presentations

Course Objectives & Outcomes

The purpose of this course is to help you build foundational skills in exploratory analytics, with an emphasis on transitioning the skills you've learned in prior classes to larger, more realistic datasets and challenges. Upon completing this course, you should expect to:

- Understand foundational skills associated with computational data exploration and manipulation, including scoping exploratory projects, engaging in data collection and wrangling, and leveraging a broad set of techniques for data sensemaking.
- Have experience with common tools in data science, including basics version control, file manipulation, program structure, and translating algorithms to code.
- Be fluent in basic algorithms and approaches for data analysis, including basic machine learning, visual analytics, and statistical analysis methods.
- Be comfortable working with data in different forms to accomplish a goal.

Assignments

- Participation (10% of your final grade): This course will require regular participation in-class. You are allowed three absences before missing class will begin to affect your grade: 3% of your participation grade will be deducted per additional absence beyond the first three. There will be no exceptions to this rule.

- Quizzes (15% of your final grade): As this course focuses classroom time primarily on questions, discussion, and in-class activities, you will be asked to complete weekly quizzes based on materials covered in the video lectures. Quizzes will be due by 11:59pm on Fridays. Note that no quizzes will be required on exam weeks. Your lowest two quiz grades will be dropped from your grade.

- Exams (25% of your final grade): You will complete two midterm exams as part of this class: one covering materials from Weeks 1-6 and a second from Weeks 7-10. Please let Dr. Szafir know by the end of Week 3 if you will need to miss either exam. Later accomodations cannot be guaranteed. Note that the entire course period before each exam will be dedicated to review.

- Homework (25% of your final grade): Each week (except for exam weeks), you will receive a set of problems that we will work on in small groups as part of our in-class activities. You are responsible for completing the full set of problems; however, some problem sets may require additional time to complete outside of class. The full problem set for each week is due by 11:59pm on Sunday evenings through the course GitHub. Note that each person must submit their own problem set even if working with others to complete the homework. Your lowest two homework scores will be dropped from your grade.

- Final Project (25% of your final grade): Over the course of the term, you will identify a stakeholder with a problem that could be addressed through data, collect and analyze data of relevance to their problem, and recommend a solution to that problem using both a formal write-up and presentation. More details about this project will be announced during Week Two. Final projects will be presented during the end-of-term Information Science Showcase, with the specific date and time TBD.

- Extra Credit (5% boost to your final grade): Dr. Szafir will release an extra credit assignment during Week Three. If you wish to complete this assignment, please submit it by 11:59pm on Monday, 10.29.

Late Policy

All assignments and projects will be due by 11:59pm on the assigned due date. Extensions to assignment deadlines will only be granted in extreme circumstances. If you need an exception or extension to an assignment deadline for any reason, please let Dr. Szafir know as soon as possible and provide reasonable documentation as to the reason for your request. Extensions cannot be granted after the due date.

Technology Requirements

Students in this course will benefit from having a laptop or tablet available for notetaking, in-class work, homework, and presentations. If you do not have access to a laptop or tablet, please consult with me. However, students and the instructor alike are encouraged to quit mail and other applications that may be distracting; to turn off notifications and ringers; and to put nonessential equipment away.

Open Discussion & Debate

(adapted from Prof. Casey Fiesler's statement, with permission and thanks)

In the classroom, students and instructors need to feel comfortable sharing their opinions and questions openly, even when we disagree. Disagreement is expected, but must be respectful and civil at all times. Students should feel welcome to share thoughts during class discussion without any fear of being disparaged for their opinions. Like yourselves, I also have opinions, and I will attempt to surface my own biases when appropriate. These disagreements or differences of opinion will not impact grades, as long as students are respectful. I invite students to meet with me to discuss concerns and ideas about how to make our learning community a positive experience for all.

This policy extends to topics and data used in the course. We may at times tackle data problems that touch on issues on which you have a strong perspective. This course takes no explicit stances on these issues, but rather is intended to be an environment where data leads our discussions. Please feel free to reach out to me should there be any topics you'd like to discuss more deeply in class or that are not currently part of the course but you would like to see

integrated. While I cannot guarantee that every request will be accommodated, I will do my best to make sure the curriculum and topics align with the needs of all of the students in the course.

Grade Distribution

This course will use a standard, 100-point grading scale:

- >= 93.0%: A
- 90.0%–92.9%: A-
- 87.0%–89.9%: B+
- 83.0%–86.9%: B
- 80.0%–82.9%: B-
- 77.0%–79.9%: C+
- 73.0%–76.9%: C
- 70.0%–72.9%: C-
- 67.0%–69.9%: D+
- 63.0%–66.9%: D
- 60.0%–62.9%: D
- < 60.0%: F

Accommodation for Disabilities

The classroom is an environment where everyone should be presented with equal opportunities to succeed. As such, I am happy to make any reasonable accommodations necessary due to disability. If you qualify for accommodations because of a disability, please submit your accommodation letter from Disability Services to me in a timely manner so that your needs can be addressed. Disability Services determines accommodations based on documented disabilities in the academic environment. Information on requesting accommodations is located on the Disability Services website (www.colorado.edu/disabilityservices/students). Contact Disability Services at 303-492-8671 or dsinfo@colorado.edu for further assistance. If you have a temporary medical condition or injury, see Temporary Medical Conditions under the Students tab on the Disability Services website and discuss your needs with your professor. I understand that emergency situations arise quickly and unexpectedly. Please let me know as soon as possible in the event that a medical situation arises that interferes with your ability to either attend or complete assigned work in this class.

Religious Holidays

Campus policy regarding religious observances requires that faculty make every effort to deal reasonably and fairly with all students who, because of religious obligations, have conflicts with scheduled exams, assignments or required attendance. In this class, I am happy to work with you to accommodate absences due to religious observance so long as I am notified of the impending absence by the end of Week 3 (9.15). Beyond this point, absences will count towards your . See the campus policy regarding religious observances for full details.

Classroom Behavior

Students and faculty each have responsibility for maintaining an appropriate learning environment. Those who fail to adhere to such behavioral standards may be subject to discipline. Professional courtesy and sensitivity are especially important with respect to individuals and topics dealing with race, color, national origin, sex, pregnancy, age, disability, creed, religion, sexual orientation, gender identity, gender expression, veteran status, political affiliation or political philosophy. Class rosters are provided to the instructor with the student's legal name. I will gladly honor your request to address you by an alternate name or gender pronoun. Please advise me of this preference early in the semester so that I may make appropriate changes to my records. For more information, see the policies on classroom behavior and the Student Code of Conduct.

Sexual Misconduct, Discrimination, Harassment and/or Related Retaliation

The University of Colorado Boulder (CU Boulder) is committed to maintaining a positive learning, working, and living environment. CU Boulder will not tolerate acts of sexual misconduct, discrimination, harassment or related retaliation against or by any employee or student. CU's Sexual Misconduct Policy prohibits sexual assault, sexual exploitation, sexual harassment, intimate partner abuse (dating or domestic violence), stalking or related retaliation. CU Boulder's Discrimination and Harassment Policy prohibits discrimination, harassment or related retaliation based on race, color, national origin, sex, pregnancy, age, disability, creed, religion, sexual orientation, gender identity, gender expression, veteran status, political affiliation or political philosophy. Individuals who believe they have been subject to misconduct under either policy should contact the Office of Institutional Equity and Compliance (OIEC) at 303-492-2127. Information about the OIEC, the above referenced policies, and the campus resources available to assist individuals regarding sexual misconduct, discrimination, harassment or related retaliation can be found at the OIEC website.

Honor Code

All students enrolled in a University of Colorado Boulder course are responsible for knowing and adhering to the academic integrity policy. Violations of the policy may include: plagiarism, cheating, fabrication, lying, bribery, threat, unauthorized access to academic materials, clicker fraud, resubmission, and aiding academic dishonesty. All incidents of academic misconduct will be reported to the Honor Code Council (honor@colorado.edu; 303-735-2273). Students who are found responsible for violating the academic integrity policy will be subject to nonacademic sanctions from the Honor Code Council as well as academic sanctions from the faculty member. Additional information regarding the academic integrity policy can be found at the Honor Code Office website.

The first instance of academic dishonesty will result in a grade of 0 on the assignment in question. Subsequent violations will result in a failing grade for the course.