SI330
Data Manipulation

Course Information

Course Meetings and Prerequisites

This course meets virtually for lectures every Tuesday from 1:00-2:20 pm Eastern Time and will be recorded. Lab will also be held synchronously (and is not recorded) from 3:00-3:50 every Tuesday. Thursday lectures are pre-recorded (except 4/15) and will be released in the morning. The prerequisite for this course is SI206: you are expected to be familiar with Python and have a basic understanding of Git. A goal of this course is to help you level up your Git skills through repeated practice, so lecture materials and assignments will be posted to GitHub. Please create a GitHub user account to receive weekly assignments.

Contact and Communications

We'll be using Google Classroom (not Canvas) as the learning management system (LMS) and form of communication. Aside from the main “Stream” tab, you can also post and respond to questions for each lab/homework by clicking on the “Classwork” tab; the calendar with Zoom invites and course deadlines are at the top of this page as well. Relevant discussion threads for each week will be monitored regularly and you can expect a response within 24 hours (or less)!
For urgent issues or other personal matters, please feel free to contact us at the email address below. We also **strongly encourage you to set up a meeting during office hours** at least once a month just to check in on how you’re doing or to ask any content-related questions. We’re happy to just have a casual chat or discuss life/career topics as well! Our hours are listed below and are conducted online. We’ve tried including a spread of hours such as nights and weekends, but if you still have conflicts (i.e. international time zones), please let us know.

You can make appointments on the calendar. While you’re free to make changes whenever (and we get that “life happens”), it’d be really appreciated if you **schedule/cancel at least 6 hours in advance. All slots are limited to 15 minutes.** If you have an extenuating circumstance that requires more time, no worries, just shoot us an email first. **Please refrain from booking more than ~2 slots/week** if it’s very busy so that we can interact with as many people as possible!

For most questions, please email us using [si330-help@umich.edu](mailto:si330-help@umich.edu)

<table>
<thead>
<tr>
<th>Instructor: Warren Li</th>
<th>GSI: Anjali Singh</th>
</tr>
</thead>
<tbody>
<tr>
<td>Email: <a href="mailto:liwarren@umich.edu">liwarren@umich.edu</a></td>
<td>Email: <a href="mailto:singhanj@umich.edu">singhanj@umich.edu</a></td>
</tr>
<tr>
<td>Office Hours (EST):</td>
<td>Office Hours (EST):</td>
</tr>
<tr>
<td>Sunday 2-3:15 pm (Int’l Slot 10 pm)</td>
<td>Wednesday: 4:00 - 5:00 pm</td>
</tr>
<tr>
<td>Monday 11 am-12 pm</td>
<td>Thursday: 11:00 am - 12:30 pm</td>
</tr>
<tr>
<td>Thursday 1-2 pm (2-2:30 Open Q&amp;A)</td>
<td></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>IA: Alana Woloshin</th>
<th>IA: Charlie Logan</th>
</tr>
</thead>
<tbody>
<tr>
<td>Email: <a href="mailto:awoloshi@umich.edu">awoloshi@umich.edu</a></td>
<td>Email: <a href="mailto:ctlogan@umich.edu">ctlogan@umich.edu</a></td>
</tr>
<tr>
<td>Office Hours:</td>
<td>Office Hours:</td>
</tr>
<tr>
<td>By appointment</td>
<td>By appointment</td>
</tr>
</tbody>
</table>

**Content and Learning Goals**

This course offers more advanced material for BSI students that deepens knowledge of programming and software development beyond the 106/206 sequence, focusing on current computing methods and data structures for obtaining, transforming, and manipulating data. The skills covered in SI 330 will be critical for future Information Analytics and Data Science courses. **This course is a programming course**—you will need to be a programmer to succeed.

Much of the work involved in data science is not just in exploration or visualization, but in the initial data manipulation/engineering stages, e.g. gathering, transforming, cleaning, and aggregating. This is the “sewer work” of data science: it’s not glamorous but is absolutely critical! Thus, while an important goal of the course is to cover specific Python-based tools for data manipulation, it is arguably even more important to teach useful high-level ways of thinking about and solving data computing problems. **It is not enough to memorize and**
recall syntax if you do not understand the process or problem conceptually. Breaking down a complex task into intermediate stages and being able to locate and process documentation when needed is crucial for any future career in technology given the fast-moving nature of the field.

Grading

We’re trying out a sort of “gamified” approach to see if that helps with managing stress and workloads. Note that there are way more points available in the course than necessary to get the highest grade! As a result, it’s not the end of the world if you miss an assignment or lab; there is some built-in flexibility and plenty of opportunities to make up lost points. Also, grades will not be curved, but we reserve the right to adjust grade boundaries in students’ favor.

- 110 pts. Python Assessment and Syllabus Quiz
- 22 x 60 pts. = 1320 pts. Participation: Clicker Questions/Discussion Forum
- 10 x 250 pts. = 2500 pts. Lab Sections (see breakdown on pg. 5)
- 10 x 500 pts. = 5000 pts. Assignments (Individual/Group)
- 10 x 200 pts. = 2000 pts. Challenge Problems
- 11 x 120 pts. = 1320 pts. PeerWise Question Generation/Answering/Review
- 100 pts. End-of-Course Evaluation and Reflection
- 1000 pts. Midterm Conferences
- 1000 pts. Final Conferences OR 1650 pts. Project* (Individual/Group)

<table>
<thead>
<tr>
<th>Final Grade Scale Total 15000 pts.</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Note: Percentages are not exact)</td>
</tr>
<tr>
<td>○ &gt;= 11600 pts. A+ (77%)*</td>
</tr>
<tr>
<td>○ 10200 pts. A (68%)*</td>
</tr>
<tr>
<td>○ 9600 pts. A- (64%)*</td>
</tr>
<tr>
<td>○ 9000 pts. B+ (60%)*</td>
</tr>
<tr>
<td>○ 8500 pts. B (57%)</td>
</tr>
<tr>
<td>○ 8000 pts. B- (53%)</td>
</tr>
<tr>
<td>○ 7000 pts. C+ (47%)</td>
</tr>
<tr>
<td>○ 6700 pts. C (45%)</td>
</tr>
<tr>
<td>○ 6400 pts. C- (43%)</td>
</tr>
<tr>
<td>○ &lt;6400 pts. E</td>
</tr>
</tbody>
</table>

*Contingent upon satisfactory (score ≥150) attendance at ≥5 labs, attempting the midterm with a score ≥600, and creating and reviewing questions ≥3 times (w/score ≥80).

Sample path to an A:

(80/110 pts. Python Assessment/Syllabus Quiz) +
(40/60 pts. * 21 classes = 880 pts. Participation) +
(9 * 240/250 pts. = 2160 pts. Lab) +
(9 * 440/500 pts. = 3960 pts. Assignments) +
(4 * 170/200 pts. = 680 pts. Challenge Problems) +

(6 * 100/120 pts. = 600 pts. PeerWise Usage) +

(100/100 pts. End-of-Course Evaluation) +

(880/1000 pts. Midterm) + (880/1000 pts. Final) = 10,220 pts.

For an A+: Replace final for project and gain an additional 550 pts. + 3 additional challenge problems for 520 pts. + 2 additional question creations for 220 pts. + increase HW average by 20 points to add 180 pts. = 11,690 pts.

---

**Text/Materials**

---

**Textbooks**

In addition to a few supplemental resources we might supply from time to time, the main (optional) text we will loosely follow along with is:


---

**Computing Resources**

You should have a functional computer with reliable internet access and a microphone/camera (see [https://techshop.umich.edu/system-recommendations](https://techshop.umich.edu/system-recommendations)).

---

**Course Schedule**

This course is split into three units. The weekly breakdown (subject to change) is as follows:

- Unit 1: Pandas (Jan/Feb)
  - 1/19: Course Welcome, Programming Assessment, Regex (NO LAB)
  - 1/28: Introduction to Numpy and Pandas
  - 2/4: Indexing, Data Cleaning, Categorical Data
  - 2/11: Groupby and Apply
  - 2/18: Merging and Pivot Tables
  - 2/25: Times Series (Rolling Windows and Resampling)
  - 3/4: Review, Testing, Debugging, and Documentation
  - 3/11-3/12: MIDTERM CONFERENCES [45 min.]
• Unit 2: SQL (Mar)
  ○ 3/16: Designing Relational Databases (NO LAB 3/16)
• Unit 3: Big Data and Applications (Apr)
  ○ 4/1: Natural Language Processing: Map Reduce, Bag of Words, Naive Bayes
  ○ 4/8: Natural Language Processing: Sentiment Analysis
  ○ 4/15: Review/Machine Learning/Networks and Graphs (+ PRESENTATIONS)
  ○ 4/22-4/23: FINAL CONFERENCES [45 min.]

---

**Course Policies**

**Attendance/Participation**

During lab sessions, you will be working on a Jupyter notebook task in groups of 4-5, which should take about 45 minutes. For the most part, we won’t be lecturing (after all, you would have just arrived from lecture!), but rather interacting with you directly. Note that it is absolutely fine (and perfectly expected) for you to make mistakes, and you won’t be judged on accuracy! This is a time for you to try things out and learn in a casual, low-stakes, and hands-on environment while being able to socialize with your peers. At the end of each class, you’ll be asked to respond to an “exit slip” for feedback/self-reflection.

For the most part, if you’re showing up and experimenting with the concepts covered in class, you’ll get full credit. Please be punctual as I want to make sure I have enough time to speak with all of you and to answer everyone’s questions as needed!

In case you were wondering, here’s the rubric for each lab:

<table>
<thead>
<tr>
<th>Points</th>
<th>Attendance/Participation</th>
<th>Deliverable</th>
<th>Reflection Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td>0 pts.</td>
<td>You were absent or almost completely disengaged.</td>
<td>You did not submit the notebook or submitted a minimal attempt.</td>
<td>You did not fill out the exit slip.</td>
</tr>
<tr>
<td>50 pts.</td>
<td>You were late, made limited contributions within your group, or did not focus well on the task.</td>
<td>You submitted a satisfactory notebook attempt.</td>
<td>You filled out the exit slip, but it is lacking in detail.</td>
</tr>
<tr>
<td>100 pts.</td>
<td>You were on time, engaged with the content, and collaborated effectively with your small group.</td>
<td></td>
<td>You filled out the exit slip with concrete ideas and complete sentences.</td>
</tr>
</tbody>
</table>
**Late/Missing Assignments**

Homework assignments will be submitted via GitHub (i.e. pushing to a repo). Feedback will be provided there as well. **Please designate one person to submit the assignment per group!**

**There are no late assignments.** But, there’s plenty of ways to make up points if you want!

**Technology Contingency Plans**

Generally speaking, there won’t be makeup labs, but please talk to me as soon as possible if you come across any extenuating circumstances; we can work things out together. Also, while it would be great to see everyone’s facial expressions in virtual class, it’s understandable if there are times where you are uncomfortable sharing your home space or have unreliable internet access. There are many ways to engage such as joining by phone. Your group mates should also strongly consider having some means of texting/messaging each other in case someone needs a low-bandwidth communication fallback.

---

**University Policies**

---

**Accommodations**

If you think you need an accommodation for a disability, please let me know at your earliest convenience. Some aspects of this course, the assignments, the in-class activities, and the way the course is usually taught may be modified to facilitate your participation and progress. As soon as you make me aware of your needs, we can work with the Services for Students with Disabilities (SSD) office to help us determine appropriate academic accommodations. SSD (734-763-3000; [http://ssd.umich.edu](http://ssd.umich.edu)) typically recommends accommodations through a Verified Individualized Services and Accommodations (VISA) form. Any information you provide is private and confidential and will be treated as such.

**Academic Integrity and Misconduct**

The University of Michigan community functions best when its members treat one another with honesty, fairness, respect, and trust. The college promotes the assumption of personal responsibility and integrity, and prohibits all forms of academic dishonesty and misconduct. All cases of academic misconduct will be referred to the LSA Office of the Assistant Dean for Undergraduate Education. Being found responsible for academic misconduct will usually result in a grade sanction, in addition to any sanction from the college. For more information, including examples of behaviors that are considered academic misconduct and potential
sanctions, please see [http://lsa.umich.edu/lsa/academics/academic-integrity.html](http://lsa.umich.edu/lsa/academics/academic-integrity.html). Also see [https://integrity.mit.edu/handbook/writing-code](https://integrity.mit.edu/handbook/writing-code) for examples of code plagiarism.

**Mental Health and Wellbeing**

The University of Michigan is committed to advancing the mental health and wellbeing of its students. If you or someone you know is feeling overwhelmed, depressed, and/or in need of support, services are available. For help, contact Counseling and Psychological Services (CAPS) at 734.764.8312 and caps.umich.edu during and after hours, on weekends and holidays, or through its counselors physically located in schools on both North and Central Campus. You may also consult University Health Service (UHS) at 734.764.8320 and [https://www.uhs.umich.edu/mentalhealthsvcs](https://www.uhs.umich.edu/mentalhealthsvcs), or for alcohol or drug concerns, see [https://www.uhs.umich.edu/aodresources](https://www.uhs.umich.edu/aodresources). For a listing of other mental health resources available on and off campus, visit [http://umich.edu/~mhealth/](http://umich.edu/~mhealth/)

**Sexual Misconduct Policy**

Title IX prohibits discrimination on the basis of sex, which includes sexual misconduct — including harassment, domestic and dating violence, sexual assault, and stalking. We understand that sexual violence can undermine students’ academic success and we encourage anyone dealing with sexual misconduct to talk to someone about their experience, so they can get the support they need. Confidential support and academic advocacy can be found with the Sexual Assault Prevention and Awareness Center (SAPAC) on their 24-hour crisis line, 734.936.3333 and at [http://sapac.umich.edu](http://sapac.umich.edu).

Alleged violations can be non-confidentially reported to the Office for Institutional Equity (OIE) at institutional.equity@umich.edu.

**COVID-19 Stipulations: Classroom Culture of Care**

UMSI is committed to delivering our mission while aiming to protect the health and safety of the community, which includes minimizing the spread of COVID-19. Our entire community is responsible for protecting the collective health of all members by being mindful and respectful in carrying out the guidelines laid out in our [Wolverine Culture of Care](http://sapac.umich.edu) and the [University’s Face Covering Policy for COVID-19](http://sapac.umich.edu). Individuals seeking to request an accommodation related to the face covering requirement under the Americans with Disabilities Act should contact the [Office for Institutional Equity](http://sapac.umich.edu).

In our classrooms all students are expected to adhere to the required safety measures and guidelines of the State of Michigan and the University of Michigan, including sanitizing their work areas, maintaining 6 feet or more of personal distance, wearing a face covering that covers
the mouth and nose in all public spaces, and not coming to class when ill or in quarantine. This course will also limit group gatherings while being thoughtful about classroom activities and exercises that require collaboration. Students who do not adhere to these safety measures while in a face-to-face class setting, and do not have an approved exception or accommodation, may be asked to disenroll from the class. For additional information refer to the OSCR Addendum to the Statement of Student Rights and Responsibilities on the OSCR website.

Final Thoughts

Given the massive amounts of readily-accessible information, it’s more important than ever to be able to organize and draw conclusions from various sources. Knowing how to manipulate data you’ve collected is a key prerequisite in the data science pipeline to help you effectively communicate a story for your audience backed by solid evidence and analysis. We hope that this course will give you practical, hands-on experience to work on projects in various applications while supplying you with the tools to independently explore past the scope here. However, times are turbulent to say the least, and while there’s been some time to adjust, it’s unavoidable that there will continue to be speed bumps along the ride (i.e. this class). With that said, we ask that you remain adaptable, and we’ll do our best to innovate and listen to your feedback to make sure this experience is valuable and engaging. We’re looking forward to seeing you (safely), and please don’t hesitate to reach out if you have any questions or concerns!