Syllabus in a Post-Covid World
SI 330
Data Manipulation

Instructional Team
Instructor: Chris Quarles
cquarles@umich.edu
GSI: Jordan Hemingway
jcheming@umich.edu
Instructional Aide: Jaik Prasad
jaiki@umich.edu

Office Hours
During our office hours, we’ll be initially offline. If you want to have video office hours, first send us a Hangouts message at cquarles or jcheming during our office hours. We’ll meet you on the Bluejeans page. We can text over Hangouts too, if you prefer.

Monday 9:30 - 11:00 or by appointment Online Chris
Tuesday 1:00 - 2:30 Online Jordan
Tuesday 3:30 - 4:30 Online Jordan
Wednesday 3:30 - 4:30 Online Jordan
Thursday 1:00 - 2:30 Online Chris

Course Description

xkcd.com/2054
Data analysis is crucial to application evaluation, as well as understanding users’ information needs. When the data required are numerous we need an automated way to gather, parse, and summarize the data. In this course, you will learn to use Python and its modules to accomplish these tasks.

**Learning Goals**

This course focuses on the following:

- Gain experience working with real world datasets
- Understand the exploration-based workflow often involved in data analysis, including the use of Jupyter Notebooks
- Understand important concepts for data manipulation (vectorization, apply(), and split-apply-combine, pivot, merge)
- Gain experience with the numpy and pandas packages in Python
- Be able to clean data, for instance by dealing with noisy and missing data
- Understand the impact data cleaning has on analysis
- Learn tools for processing text data including the use of regular expressions
- Be aware of - and get experience dealing with - different types of data
- Work with a variety of data storage formats (csv, json, etc.)
- Work with a relational database (SQL)
- Learn methods for working with big data (MapReduce, Spark)

**Assessments**

<table>
<thead>
<tr>
<th>Assessment</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>In-class notebooks</td>
<td>10%</td>
</tr>
<tr>
<td>(Due two days after class at 1pm)</td>
<td></td>
</tr>
<tr>
<td>Homework</td>
<td>30%</td>
</tr>
<tr>
<td>Due when it’s due</td>
<td></td>
</tr>
<tr>
<td>In-class midterm exams (2)</td>
<td>40%</td>
</tr>
<tr>
<td>Correction: The second exam will be a take-home exam.</td>
<td></td>
</tr>
<tr>
<td>Final project</td>
<td>20%</td>
</tr>
</tbody>
</table>

**Textbook**

We will have readings in this course, which will all be free to you to read online. The first text will be Python for Data Analysis by Wes McKinney. You can read it online here: [https://www.oreilly.com/library/view/python-for-data/9781491957653](https://www.oreilly.com/library/view/python-for-data/9781491957653)

**Grades**

The table below determines your grade, with +’s and -’s interspersed evenly at intervals of 3.33%. I reserve the right to curve the grading scale in the class’s favor.

<table>
<thead>
<tr>
<th>Percentage</th>
<th>Minimum Grade</th>
</tr>
</thead>
<tbody>
<tr>
<td>93.3</td>
<td>A</td>
</tr>
<tr>
<td>83.3</td>
<td>B</td>
</tr>
<tr>
<td>73.3</td>
<td>C</td>
</tr>
<tr>
<td>63.3</td>
<td>D</td>
</tr>
</tbody>
</table>

**Late Assignments**
Life happens. We get it. Life happens to us too. To accommodate the demands of life, you get three
late days. These days can be applied to anything but exams. You can turn one assignment in 3 days late, 3
assignments in one day late, etc. You don’t need to tell us how you want to allocate your late days. Just turn
your assignment in. If you turn your assignments in collectively more than 3 days late, we will assign your
late days in the way that benefits you the most. Any late assignments beyond the 3 late days will be treated
as zeroes, unless arrangements are made due to something like a serious illness. All late days will be
rounded up to the nearest day. For example, an assignment that gets turned in 15 minutes after the deadline
will be considered 1 day late. These calculations will be done at the end of the semester. Until then, all
assignments will look like they count toward your grade in Canvas. So, if you have a lot of late
assignments, expect your Canvas-reported grade to drop a bit at the end of the semester.

If you need to miss class or assignments because of a serious injury, illness, death in the family, UM
sports competition, etc., we can definitely make arrangements. Please contact me as early as is reasonable
for you, so that we can minimize the impact to your learning.

Attendance
Come to class every day. I’ll do my best to make it worth your time. You’re spending a lot of money on
this college education. Get what you can out of it. Also, we’ll be doing things that affect your grade.

Communication
For general outside-of-class communication, we’ll be using Slack. Slack is an app that allows teams to
communicate. This allows (a) you to see the questions other students are asking, (b) you to answer the
questions other students are asking, and (c) opportunities for better collaboration with classmates. If you
have homework questions, or questions about the class in general, it’s best to post them to Slack. If you
have more private questions about, for example, an illness, then you should probably email me. If you have
questions about how your homework assignment was graded, start with Jordan. These are just general
guidelines. You’re always welcome to send me an email. (The link to join the Slack workspace can be
found on the home page in Canvas.)

Unless there’s a potentially urgent concern, please mute and put away all cell phones. Also, turn off all
messaging notifications on your computer. Those constant distractions decrease learning significantly. If
we see you texting or chatting during class, we will politely ask you to put your device away or turn off
notifications.

Announcements will be over Canvas.

Accommodations for Students with Disabilities
If you have or think you have a disability which may affect your ability to do well in this class, you should
talk to me soon. That way we can work with the Office of Services for Students with Disabilities (SSD) to
make sure you get the appropriate accommodations. You can contact SSD at 734-763-3000 or
http://www.umich.edu/sswd/ They typically recommend accommodations through a Verified
Individualized Services and Accommodations (VISA) form. We will treat any information you provide as
private and confidential.

Plagiarism 101: What is “Original Work?”
When I’m using code to analyze data, I look stuff up online. A LOT. That’s expected - and necessary.
There are some tricky problems that are much easier to find on StackExchange than in some programmer’s
(badly written) documentation. In a class, our goal is to learn. Looking things up or asking friends for help
is part of that learning process. However, your code needs to be your own. Copy-paste isn’t learning. It’s
not only plagiarism, it’s a bad use of your time at this prestigious institution.
What’s OK:
- Asking a friend what command they used to do a particular task. (Such as .split() or join())
- Discussing general strategies for problem solving (after you’ve thought about it yourself)
- Doing an internet search for how to solve problems or resolve errors, such as “how to convert dict to pandas dataframe”
- Copying a few lines of code from StackExchange, making sure to attribute the source in your code comment

What’s Not OK:
- Copying your friend’s code
- Copying your friend’s code, but changing the variable names
- Copying your friend’s code, but typing it yourself and adding little things in to confuse the grader
- Copying large blocks of code from online, especially without attribution

What Will Happen:
- The first instance of clear plagiarism will result in a zero on that assignment
- The second instance of plagiarism will result in a failing grade in this course.

Course Summary:
- [https://umich.instructure.com/courses/343764/assignments/syllabus](https://umich.instructure.com/courses/343764/assignments/syllabus)

<table>
<thead>
<tr>
<th>Date</th>
<th>Details</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fri Jan 10, 2020</td>
<td>Assignment <a href="#">In-Class Notebook 1</a> due by 1pm</td>
</tr>
<tr>
<td>Wed Jan 15, 2020</td>
<td>Assignment <a href="#">In-Class Notebook 2</a> due by 1pm</td>
</tr>
<tr>
<td>Thu Jan 16, 2020</td>
<td>Assignment <a href="#">Homework 1: Py4DA Ch. 4 and 5 Quiz</a> due by 1pm</td>
</tr>
<tr>
<td>Fri Jan 17, 2020</td>
<td>Assignment <a href="#">In-Class Notebook 3</a> due by 1pm</td>
</tr>
<tr>
<td>Wed Jan 22, 2020</td>
<td>Assignment <a href="#">In-Class Notebook 4</a> due by 1pm</td>
</tr>
<tr>
<td>Thu Jan 23, 2020</td>
<td>Assignment <a href="#">Homework 2: Vectorization and Filtering</a> due by 1pm</td>
</tr>
</tbody>
</table>
Fri Jan 24, 2020
Assignment  In-Class Notebook 5  due by 1pm

Wed Jan 29, 2020
Assignment  In-Class Notebook 6  due by 1pm

Thu Jan 30, 2020
Assignment  Homework 3  due by 1pm

Fri Jan 31, 2020
Assignment  In-Class Notebook 7  due by 1pm

Wed Feb 5, 2020
Assignment  In-Class Notebook 8  due by 1pm

Fri Feb 7, 2020
Assignment  In-Class Notebook 9  due by 1pm

Sat Feb 8, 2020
Assignment  Homework 4  due by 1pm

Thu Feb 13, 2020
Assignment  Midterm 1  due by 2:30pm

Wed Feb 19, 2020
Assignment  In-Class Notebook 10  due by 1pm

Fri Feb 21, 2020
Assignment  In-Class Notebook 11  due by 1pm

Thu Feb 27, 2020
Assignment  Homework 6  due by 1pm
Fri Feb 28, 2020
Assignment In-Class Notebook 12 due by 1pm

Wed Mar 11, 2020
Assignment In-Class Notebook 13 due by 1pm

Thu Mar 12, 2020
Assignment Homework 7 due by 1pm

Thu Mar 19, 2020
Assignment In-Class Notebook 14 due by 1pm

Sat Mar 21, 2020
Assignment In-Class Notebook 15 due by 1pm

Sun Mar 22, 2020
Assignment Peer Reviews due by 11:59pm

Tue Mar 24, 2020
Assignment Project Work Plan due by 11:59pm

Thu Mar 26, 2020
Assignment In-Class Notebook 16 due by 11:59pm

Fri Mar 27, 2020
Assignment Homework 8 due by 12pm

Thu Apr 2, 2020
Assignment Midterm 2, Part 1 due by 1pm
Assignment Midterm 2, Part 2 due by 1pm
Assignment Midterm 2, Part 3 due by 1pm
Tue Apr 7, 2020

Assignment  In-Class Lesson 17  due by 11:59pm

Thu Apr 9, 2020

Assignment  In-Class Notebook 18  due by 1pm

Assignment  Midterm 1
(3 students)

Assignment  Unnamed Quiz

April 2020

Calendar

<table>
<thead>
<tr>
<th>Sunday</th>
<th>Monday</th>
<th>Tuesday</th>
<th>Wednesday</th>
<th>Thursday</th>
<th>Friday</th>
<th>Saturday</th>
</tr>
</thead>
<tbody>
<tr>
<td>29 Previous</td>
<td>30 Previous</td>
<td>31 Previous</td>
<td>1</td>
<td>2 Click to view</td>
<td>3</td>
<td>4</td>
</tr>
<tr>
<td>month</td>
<td>month</td>
<td>month</td>
<td></td>
<td>event details</td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>7</td>
<td>8 Today</td>
<td>9 Click to view</td>
<td>10</td>
<td>11</td>
</tr>
<tr>
<td></td>
<td></td>
<td>Click to view</td>
<td></td>
<td>event details</td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>13</td>
<td>14</td>
<td>15</td>
<td>16</td>
<td>17</td>
<td>18</td>
</tr>
<tr>
<td>19</td>
<td>20</td>
<td>21</td>
<td>22</td>
<td>23</td>
<td>24</td>
<td>25</td>
</tr>
<tr>
<td>26</td>
<td>27</td>
<td>28</td>
<td>29</td>
<td>30</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3 Next month</td>
<td>4 Next month</td>
<td>5 Next month</td>
<td>6 Next month</td>
<td>7 Next month</td>
<td>8 Next</td>
<td>9 Next</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>month</td>
<td>month</td>
</tr>
</tbody>
</table>

Assignments are weighted by group:

<table>
<thead>
<tr>
<th>Group</th>
<th>Weight</th>
</tr>
</thead>
<tbody>
<tr>
<td>In-Class Notebooks</td>
<td>10%</td>
</tr>
<tr>
<td>Homework</td>
<td>30%</td>
</tr>
<tr>
<td>Midterms</td>
<td>40%</td>
</tr>
<tr>
<td>Project</td>
<td>20%</td>
</tr>
<tr>
<td>---------</td>
<td>-----</td>
</tr>
<tr>
<td>Total</td>
<td>100%</td>
</tr>
</tbody>
</table>