

Course Syllabus for SIADS 505: Data Manipulation

Course Overview and Prerequisites

This course will introduce data manipulation and cleaning techniques using the popular python pandas data science library and introduce the abstraction of the Series and DataFrame as the central data structures for data analysis, along with an understanding of how to use functions such as groupby, merge, and pivot tables effectively. By the end of this course, students will be able to take tabular data, clean it, manipulate it, and run basic inferential statistical analyses.

There are no course prerequisites.

Instructor and Course Assistants

Instructor: Chris Brooks (broosch@umich.edu)

Course Assistant: Zihan Wu (ziwu@umich.edu);

Anthony Giove (agiove@umich.edu)

Greg Myers (gamyers@umich.edu)

Communication Expectations

Contacting instructor and course assistant: Please only use the course channel in Slack. Please ask all questions in public if possible so others can learn from your question. If you need to ask a private question please direct message both the instructor and the course assistant.

Email response time: 48 hours

Slack response time: 48 hours

Office hours: as noted in the coursera platform

Office hours will be held weekly and all office hours will have the same passcode **505**.

Live Office Hours (EDT): August 30 to September 28

Monday 1-2pm (Chris - Weekly Introduction)
Tuesday 9-10am (Anthony - Assignment Review)
Wednesday 1-2pm (Greg - Assignment)
Thursday 7pm-8pm (Zihan - Extra Topics)
Friday 8-9pm (Zihan - Assignment)
Saturday 2-3pm (Greg - Assignment)
Saturday 9-10pm (Extra Topics)
Sunday 1-2pm (Greg - Assignment)

Required Textbook (if relevant)

[Python for Data Analysis: Data Wrangling with Pandas, NumPy, and iPython](#) 2nd edition by Wes McKinney (O'Reilly). Copyright 2017 Wes McKinney, 978-1-491-95766-0.

Textbook link provided above allows free usage through the University of Michigan Library. University credentials required.

Learning Outcomes

1. Create and use pandas DataFrames to represent raw data
2. Extract information from text data using regular expressions
3. Understand high level patterns of data cleaning and how they are realized in python
4. Understand how other formats embed semantics in data and how one can approach representing them in tabular formats
5. Understand the issues of replication in data science and how it starts with data manipulation
6. Have an awareness of other kinds of structured data such as networks, graphs, natural language

Course Schedule

This course **begins on Monday, August 30, 2021**, and **ends on Sunday, September 26, 2021**.

Assignments will be due as noted in the Coursera platform (roughly weekly, on Sunday's at 11:59pm EST).

Grading

Course Assignments	Percentage of Final Grade
Week 1 Jupyter Notebook Assignment	25% (total)
Week 2 Jupyter Notebook Assignment	25% (total)
Week 3 Jupyter Notebook Assignment	25%
Week 4 Jupyter Notebook Assignment	25%

All assignments can be resubmitted **as many times as you like** up until the **assignment deadline** with the best submission count towards the learner grade.

Letter Grades, Course Grades, and Late Submission Policy

Refer to the [MADS Assignment Submission and Grading Policies](#) section of the UMSI Student Handbook (access to Student Orientation course required)

For this course, no late assignments will be accepted (0%). Extenuating circumstances will be considered (please reach out to umsimadshelp@umich.edu as soon as possible).

Percentage grades will be converted to letter grades using the following formula (Note that an A+ is only awarded by doing a bonus!):

Total Assignment Scores	Letter Equivalent
>100	A+
95-100	A
90-95	A-
85-90	B+

80-85	B
75-80	B-
70-75	C+
65-70	C
60-65	C-
55-60	D+
50-55	D
<50	E

Accommodations

Refer to the [Accommodations for Students with Disabilities](#) section of the UMSI Student Handbook (access to the Student Orientation course required).

Use the [Student Intake Form](#) to begin the process of working with the University's Office of Services for Students with Disabilities.

Accessibility

[Screen reader configuration for Jupyter Notebook Content](#)

Help Desk(s): How to get help

- Degree program questions or general help - umsimadshelp@umich.edu
- Coursera's Technical Support (24/7) - <https://learner.coursera.help/>

Library Access

Refer to the [U-M Library's information sheet](#) on accessing library resources from off-campus. For more information regarding library support services, please refer to the [U-M Library Resources](#) section of the UMSI Student Handbook (access to the Student Orientation course required).

Student Mental Health

Refer to the University's [Resources for Stress and Mental Health website](#) for a listing of resources for students.

Student Services

Refer to the [Introduction to UMSI Student Life](#) section of the UMSI Student Handbook (access to the Student Orientation course required).