

GRADTDA 5622 - Big Data Computing Foundations II

Syllabus

Course Information

- Course times: Asynchronous
 - Weekly synchronous meeting (optional): TBD
- Credit hours: 3
- Mode of delivery: Distance Learning

Instructor

- Name: Thomas Bihari
- Email: bihari.5@osu.edu
- Office location: Zoom - See Syllabus page on CarmenCanvas for link.
- Office hours: See Syllabus page on CarmenCanvas for link.
- Preferred means of communication:
 - My preferred method of one-to-one communication is email.
 - I will send class-wide communications via the Announcements tool in CarmenCanvas. Please check your [notification preferences](#) to ensure you are notified of new announcements.
 - Class-wide discussions will be in Microsoft Teams. See Syllabus page on CarmenCanvas for link.

Grader

- Name: TBD
- Email: TBD
- Office location: TBD
- Office Hours: TBD
- Preferred means of communication: TBD

Course Prerequisites

GRADTDA 5621. Good standing in the MTDA program. Course enrollment is restricted to graduate students enrolled in the program. Some experience in Data Analysis and suitable undergraduate preparation is beneficial.

Course Description

The successful working professional engaged in real-world, problem-solving contexts often found in many modern enterprises must be able to construct schemas that locate data sources and scrape them, followed by practical workflows that ingest and clean data, extract useful information for exploration and visualization, and use that information to address work-related challenges. A successful outcome of this instructional offering is to

construct programming-based data-driven tools to facilitate context-aware problem-solving. The enrollee will also gain proficiency in identifying, sourcing, manipulating, and interpreting data. Further, he or she will also be highly capable of creating workflows for a variety of data types and representations.

This is the second course in a two-semester sequence comprised of two, 3 credit-hour courses focusing on the use of Python programming and tools/environments. The sequence is to be taken in parallel with the sequence focusing on fundamental statistical data analytic methods.

Learning Outcomes

By the end of this course, students should be able to:

- Use the programming language of Python and the companion integrated development environment of Jupyter to import, manipulate and visualize data. Learn to interface with other environments and open-source software libraries and tools.
- Understand the process and flow of data-driven problem solving.
- Understand the challenges of working with large datasets, the drivers of complexity, and the tools and techniques for measuring time and space performance.
- Understand the tools for addressing complexity. Mastery of data organization and query optimization techniques.
- Achieve competency with distributed algorithms and environments such as Map-reduce, Spark, and/or Hadoop.
- Gain exposure to current cloud-based data management technologies.
- Learn to create high-utility and efficacious workflows using Python-based tools to extract information inherent in the data from a variety of applications.
- Create high-information and useful tools that interface with the data analytic workflows to create meaningful and useful visualizations and in turn realize for the use of feedback loops to alter the data analysis in useful ways.
- Conduct application-driven, exploratory analyses that use other contextual information to contribute to problem solving and achieve true translation.
- Author effective summary reports of the performed analysis using Jupyter notebooks and/or slide presentations.

Dos and Don'ts

Mode of delivery: This course is 100% online. There are no required instructional sessions when you must be logged in to Carmen at a scheduled time. The only exceptions for attendance will be quizzes, exams, and presentations.

The pace of online activities: This course is divided into modules that are released days ahead of time. Students are expected to keep pace with deadlines but may schedule their efforts freely within that timeframe.

Credit hours and work expectations: This is a 3 credit-hour course. According to [Ohio State bylaws on instruction](#), students should expect around 3 hours per week of time spent on direct instruction (instructor content and Carmen activities, for example) in addition to 6 hours of homework (reading and assignment preparation, for example) to receive a grade of C average. Computing courses often require up to 12-15 hours of work.

Attendance and participation requirements: Because this is an online course, your attendance is based on your online activity and participation. The following is a summary of students' expected participation:

- Participating in online activities for attendance: at least once per week
 - You are expected to log in to the course in Carmen/Teams every week. During most weeks you will probably log in many times. If you have a situation that might cause you to miss an entire week of interaction, discuss it with me as soon as possible.
- Participating in discussion forums: two or more times per week

- As part of your participation, each week you can expect to post at least twice as part of our substantive class discussion on the week's topics.
- Weekly retrospective
 - Each week, you will fill out a very short retrospective. You get full credit simply by filling out the retrospective.
- Knowledge checks
 - Some topics will include a simple knowledge checks, to help you assess your understanding of the topic. You get full credit simply by taking the quiz.
- Office hours and live sessions: optional
 - All live, scheduled events for the course, including my office hours, are optional. You are not graded on attendance on these events. The exceptions are the days of quizzes, exams, and presentations.

Course Materials and Technologies

Required Texts (for now)

- A. **Python for Data Analysis - Data Wrangling with Pandas, NumPy, and IPython**, William McKinney, O'Reilly Media, 2021, Third edition.
 - a. An electronic version is available for online reading [on the OSU Safari website](#). You will need to be logged in to OSU Safari (O'Reilly Media).
- B. **Python Data Science Handbook** by Jake VanderPlas. Copyright © 2017 Jake VanderPlas.
 - a. An electronic version is available for online reading [on the OSU Safari website](#). You will need to be logged in to OSU Safari (O'Reilly Media).
- C. **Learning Spark**, 2nd Edition by Jules S. Damji, Brooke Wenig, Tathagata Das, Denny Lee. Released July 2020. O'Reilly Media, Inc. ISBN: 9781492050049
 - a. An electronic version is available for online reading [on the OSU Safari website](#). You will need to be logged in to OSU Safari (O'Reilly Media).
- D. **Mining of Massive Datasets**, 3rd Edition by Jure Leskovec, Anand Rajaraman, Jeff Ullman. A free electronic version is available here: <http://mmds.org/>.

Required Equipment

- **Computer:** current Mac (macOS), PC (Windows 10+, and Linux) with high-speed internet connection
- **Webcam:** built-in or external webcam, fully installed and tested
- **Microphone:** built-in laptop or tablet mic or external microphone
- **Other:** a mobile device (smartphone or tablet) to use for BuckeyePass authentication

Required Software

- Microsoft Office 365: All Ohio State students are now eligible for free Microsoft Office 365. Visit the [installing Office 365 link](#).
- We will be using the Ohio Supercomputer Center's Python / Jupyter environments for most assignments (to be discussed in class). Optionally, you might want to use other Python / Jupyter environments, such

as:

- [Anaconda](#) (Python / Jupyter environment runs on PCs, Macs, etc.)
- [Google Collab](#)

CarmenCanvas Access

You will need to use [BuckeyePass](#) multi-factor authentication to access your courses in Carmen. To ensure that you are able to connect to Carmen at all times, it is recommended that you do each of the following:

- Register multiple devices in case something happens to your primary device. Visit the [BuckeyePass - Adding a Device](#) help article for step-by-step instructions.
- Request passcodes to keep as a backup authentication option. When you see the Duo login screen on your computer, click Enter a Passcode and then click the Text me new codes button that appears. This will text you ten passcodes good for 365 days that can each be used once.
- [Install the Duo Mobile application](#) on all of your registered devices for the ability to generate one-time codes in the event that you lose cell, data, or Wi-Fi service.

If none of these options will meet the needs of your situation, you can contact the IT Service Desk at 614-688-4357 (HELP) and IT support staff will work out a solution with you.

Technology Skills Needed for This Course

- [Navigating CarmenCanvas](#)
- [CarmenZoom virtual meetings](#)
- [Ohio Supercomputer Center](#) resources (will be covered in class)

Technology Support

For help with your password, university email, CarmenCanvas, or any other technology issues, questions or requests, contact the IT Service Desk, which offers 24-hour support, seven days a week.

- Self Service and Chat: osu.edu/it
- Phone: 614-688-4357 (HELP)
- Email: servicedesk@osu.edu

Digital Flagship

Digital Flagship is a student success initiative aimed at helping you build digital skills for both college and your career. This includes offering an engaging collection of digital tools and supportive learning experiences, university-wide opportunities to learn to code, and a Design Lab to explore digital design and app development. Digital Flagship resources are available to help Ohio State students include on-demand tutorials, The Digital Flagship Handbook (your guide for all things tech-related), workshops and events, one-on-one tech consultations with a peer or Digital Flagship staff member, and more. To learn more about how Digital Flagship can help you use technology in your courses and grow your digital skills, [visit this link to their website](#).

Grading and Faculty Response

How Your Grade is Calculated

Assignment Category	Points
Participation (activity, retrospectives, discussions, practice quizzes – see above)	10%
Individual Homeworks (5 x 6%) – content	30%
Groupwise Case Studies (4 x 10%) – content, participation	40%
Group Final Project – content, participation	20%

Academic integrity and collaboration:

Your written assignments, including discussion posts, should be your own original work. In formal assignments, you should follow the MLA/APA/Chicago, etc. style to cite the ideas and words of your research sources. When you collaborate in teams, it is assumed that you will contribute diligently and sincerely. All team members should agree to their roles before the start of all exercises. In case of conflicts, the instructional staff should be notified. This is essential also for fairness in grading.

Submission of Assignments

No assignment will be deemed late until the end-of-semester. There will be recommended deadlines. It will be recommended that all homework be submitted by the suggested dates. The goal is to complete and not adhering to strict deadlines. This is a shoutout to all those who work, have other obligations, yet chose to enroll in this class.

Instructor(s) Feedback and Response Time

Remember that you can call 614-688-4357 (HELP) at any time if you have a technical problem.

- **Preferred contact method:** If you have a one-on-one question, please contact me first through Ohio State email address. I will reply to emails within two days when class is in session at the university, typically sooner.
- **Class announcements:** We will send all important class-wide messages through the Announcements tool in CarmenCanvas. Please check [your notification preferences](#), to ensure you receive these messages.
- **Discussion board:** I will reply to discussion board posts within two days when class is in session at the university, typically sooner.
- **Grading and feedback:** For large assignments, you can generally expect feedback within two weeks, typically much sooner.

Grading Scale

93–100: A
90–92.9: A-
87–89.9: B+
83–86.9: B
80–82.9: B-
77–79.9: C+
73–76.9: C
70–72.9: C-
67–69.9: D+
60–66.9: D
Below 60: E

Other Course Policies

Discussion and Communication Guidelines

The following are my expectations for how we should communicate as a class. Above all, please remember to be respectful and thoughtful.

- **Writing style:** While there is no need to participate in class discussions as if you were writing a research paper, you should remember to write using good grammar, spelling, and punctuation. A more conversational tone is fine for non-academic topics.
- **Tone and civility:** Let us maintain a supportive learning community where everyone feels safe and where people can disagree amicably. Remember that sarcasm doesn't always come across online. Also, please be cognizant of the wide variety of backgrounds in the class.
- **Citing your sources:** When we have academic discussions for case studies and projects, please cite your sources to back up what you say. For the textbook or other course materials, list at least the title and page numbers. For online sources, include a link. For manuscripts use appropriate and suitable citation style (see above).
- **Backing up your work:** Consider composing your academic posts in a word processor, where you can save your work, and then copying into the Carmen discussion.

Academic Integrity Policy

See [Descriptions of Major Course Assignments](#) for specific guidelines about collaboration and academic integrity in the context of this online class.

Ohio State's Academic Integrity Policy

Academic integrity is essential to maintaining an environment that fosters excellence in teaching, research, and other educational and scholarly activities. Thus, The Ohio State University and the Committee on Academic Misconduct (COAM) expect that all students have read and understood the university's [Code of Student Conduct](#), and that all students will complete all academic and scholarly assignments with fairness and honesty.

Students must recognize that failure to follow the rules and guidelines established in the university's Code of Student Conduct and this syllabus may constitute "Academic Misconduct." The Ohio State University's Code of

Student Conduct (Section 3335-23-04) defines academic misconduct as: “Any activity that tends to compromise the academic integrity of the university or subvert the educational process.” Examples of academic misconduct include (but are not limited to) plagiarism, collusion (unauthorized collaboration), copying the work of another student, and possession of unauthorized materials during an examination. Ignorance of the university’s Code of Student Conduct is never considered an excuse for academic misconduct, so I recommend that you review the Code of Student Conduct and, specifically, the sections dealing with academic misconduct.

If we suspect that a student has committed academic misconduct in this course, I am obligated by university rules to report my suspicions to the Committee on Academic Misconduct. If COAM determines that you have violated the university’s Code of Student Conduct (i.e., committed academic misconduct), the sanctions for the misconduct could include a failing grade in this course and suspension or dismissal from the university. If you have any questions about the above policy or what constitutes academic misconduct in this course, please contact me. Other sources of information on academic misconduct (integrity) to which you can refer include:

- [Committee on Academic Misconduct](#)
- [Ten Suggestions for Preserving Academic Integrity](#)
- [Eight Cardinal Rules of Academic Integrity](#)

Copyright for Instructional Materials

The materials used in connection with this course may be subject to copyright protection and are only for the use of students officially enrolled in the course for the educational purposes associated with the course. Copyright law must be considered before copying, retaining, or disseminating materials outside of the course.

Statement on Title IX

All students and employees at Ohio State have the right to work and learn in an environment free from harassment and discrimination based on sex or gender, and the university can arrange interim measures, provide support resources, and explain investigation options, including referral to confidential resources. If you or someone you know has been harassed or discriminated against based on your sex or gender, including sexual harassment, sexual assault, relationship violence, stalking, or sexual exploitation, you may find information about your rights and options on [Ohio State’s Title IX website](#) or by contacting the Ohio State Title IX Coordinator at titleix@osu.edu. Title IX is part of the Office of Institutional Equity (OIE) at Ohio State, which responds to all bias-motivated incidents of harassment and discrimination, such as race, religion, national origin, and disability. For more information, visit the [OIE website](#) or email equity@osu.edu.

Commitment to a Diverse and Inclusive Learning Environment

The Ohio State University affirms the importance and value of diversity in the student body. Our programs and curricula reflect our multicultural society and global economy and seek to provide opportunities for students to learn more about persons who are different from them. We are committed to maintaining a community that recognizes and values the inherent worth and dignity of every person; fosters sensitivity, understanding, and mutual respect among each member of our community; and encourages each individual to strive to reach their own potential. Discrimination against any individual based upon protected status, which is defined as age, color, disability, gender identity or expression, national origin, race, religion, sex, sexual orientation, or veteran status, is prohibited.

Your Mental Health

As a student you may experience a range of issues that can cause barriers to learning, such as strained relationships, increased anxiety, alcohol/drug problems, feeling down, difficulty concentrating, and/or lack of motivation. These mental health concerns or stressful events may lead to diminished academic performance or reduce a student's ability to participate in daily activities. No matter where you are engaged in distance learning, The Ohio State University's Student Life Counseling and Consultation Service (CCS) is here to support you. If you find yourself feeling isolated, anxious, or overwhelmed, [on-demand mental health resources](#) are available. You can reach an on-call counselor when CCS is closed at 614- 292-5766. 24-hour emergency help is available through the [National Suicide Prevention Lifeline website](#) or by calling 1-800-273-8255(TALK). [The Ohio State Wellness app](#) is also a great resource.

Accessibility Accommodations for Students with Disabilities

Requesting Accommodations

The university strives to make all learning experiences as accessible as possible. If you anticipate or experience academic barriers based on your disability including mental health, chronic or temporary medical conditions, please let me know immediately so that we can privately discuss options. To establish reasonable accommodations, I may request that you register with [Student Life Disability Services \(SLDS\)](#) ([Links to an external site](#)). After registration, make arrangements with me as soon as possible to discuss your accommodations so that they may be implemented in a timely fashion.

Disability Services Contact Information

- Phone: 614-292-3307
- Website: osu.edu
- Email: slds@osu.edu
- In-person: [Baker Hall 098, 113 W. 12th Avenue](#)

Accessibility of Course Technology

This online course requires the use of CarmenCanvas (Ohio State's learning management system) and other online communication and multimedia tools. If you need additional services to use these technologies, please request accommodations with your instructor.

- [CarmenCanvas accessibility](#)
- Streaming audio and video - talk to instructors
- [CarmenZoom accessibility](#)
- Collaborative course tools - talk to instructors