

# GRADTDA 5621:Big Data Computing Foundations I

## Instructor Contact Information

Name: **Raghu Machiraju**  
Office Location: **Dreese 495**  
Email: **[machiraju.1@osu.edu](mailto:machiraju.1@osu.edu)**  
Office Hours: **TBD**

## Course Coordinator Information

Name: Chaitanya Kulkarni  
Email: **[kulkarni.132@osu.edu](mailto:kulkarni.132@osu.edu)**  
Office Hours: **TBD**

## Course Meeting Information

Method: **Online**  
Course Hours: **\*\*Course is assumed to contain 42 one-hour lectures and project presentations.**

## Course Prerequisites

Good standing in the MTDA program. Course enrollment is restricted to graduate students enrolled in the program.

## Description of Course:

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

## Course Learning Goals:

**Upon successful completion of the course, students will be able to:**

1. Use the programming languages of Python and Javascript and the companion integrated development environment of Jupyter to import, manipulate and visualize data. Learn to interface with other environments such as the R programming language and open-source software libraries and tools.
2. Learn to locate data sources and scrape data from collections and ingest them. Clean the data and wrangle or munge them into formats and representations amenable for further analysis and visualization. Become adept at constructing workflows to scrape and wrangle data for a variety of applications.
3. Learn to create high-utility and efficacious workflows using Python-based tools to extract summary information inherent in the data from a variety of

applications.

4. Create high-information and useful Javascript/Java based 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.
5. Conduct application-driven, exploratory analyses that use other contextual information including ontologies to contribute to problem solving and achieve true translation.
6. Author effective summary reports of the performed analysis using Jupyter notebooks.

#### **Course Materials and Texts:**

##### **Textbooks:**

1. Python for Data Analysis - Data Wrangling with Pandas, NumPy, and IPython, William McKinney, O'Reilly Media, October 2017. (An electronic version is available for online reading through the OSU Safari website)
2. JavaScript: The Definitive Guide, 6th Edition, David Flanagan, O'Reilly Media, May 2011. (An electronic version is available for online reading through the OSU Safari website)

#### **Grading Information:**

##### **Grade Breakdown:**

Homework = 30%  
Case Studies = 40%  
Final Project = 30%

##### **Grading Scale:**

>90% A/A-  
80-90% B-/B/B+  
70-80% C-/C/C+  
60-70% D/D+  
<60% E

#### **Course Delivery:**

---

Asynchronous (pre-recorded at student's leisure): each week several lecture videos, totaling approximately 2.5 to 3 hours of lecture, will be posted on the course website. You are responsible for watching the videos and studying the material that is assigned each week. In addition to the lecture videos, weekly assignments will be posted on Carmen.

Synchronous (live): there will also be occasional group discussions or presentations over Zoom, which will be scheduled in advance with students. Office hours and meetings or discussions between individual students and myself and grader can be scheduled in advance and are optional as needed.

Overall, the majority of the course content can be completed on a student's own time at their leisure, within the constraints of deadlines for each lecture video or topic.

### **Attendance and Participation:**

---

Because this is a distance-education course, your attendance is based on your online activity and participation. Please check your email and log into Carmen at least once per weekday, and participate in discussions on Carmen at least three times per week.

### **Discussion and Communication:**

---

When posting content visible to myself and/or to other students, please remember to be respectful, thoughtful, and professional.

- Tone and civility: all discussions must contribute to a sense of safety and civility to ensure all students and myself feel comfortable and welcome as part of the course.
- Please be mindful of diversity of backgrounds and thought and strive to respect and uphold diversity.
- Writing style: please use good grammar, spelling, and punctuation. No hand written material will be accepted.
- Cite your sources: in academic discussions, please cite your sources to back up what you say. This may be as simple as links to resources you have found online, or references to page numbers in the textbook.

### **Assignments and Grades:**

---

Homework will account for 30% of your course grade. They will be assigned and submitted via Carmen. Homework assignments will usually require code reviews, algorithm summaries, data analysis, and some coding.

Case studies will account for 40% of your course grade. They will be completed in smaller groups and will entail data analysis, coding, and reporting components. They will be assigned and submitted via Carmen as Jupyter notebooks.

In lieu of final exam, a final project comprising of 30% of the course will be required. The presentations will be spread and over the last week and the final exam hour (if needed).

### **Health and Safety:**

---

The Ohio State University Wexner Medical Center's Coronavirus Outbreak site (<https://wexnermedical.osu.edu/features/coronavirus>) includes the latest information about COVID-19 as well as guidance for students, faculty and staff. Guidelines and requirements for campus safety from the University's COVID-19 Transition Task Force were published on July 1 on the Safe and Healthy website (<https://safeandhealthy.osu.edu>). They include the following:

- A daily health check to report body temperature and health status will be required for all faculty, staff and students each day they intend to be on Ohio State's campuses in the autumn.
- Face masks must be worn in indoor settings, including classrooms.
- Members of the campus community will be required to sign a pledge "to affirm their understanding of what is needed to help fight the spread of the virus and their intention to do their part."
- Accountability measures will be in place for those who refuse to abide by required health and safety guidelines.

You are encouraged to complete daily health checks and to self-isolate if running a fever or are in other ways symptomatic.

### Potential Disruptions to Education:

---

If you are to complete content or complete assignments because of positive diagnosis or symptoms, please reach out to me as soon as possible to let me know so we can develop a contingency plan. Typically, this will include postponing due dates or developing alternative assignments, and will be done on a case by case basis.

If I am unable to create and post new content due to positive diagnosis or symptoms, I will alert the class as quickly as possible to develop a contingency plan.

### Lecture Breakdown:

Lecture	Topic
Week 1	Introduction and case-studies of workflows from various applications and using data of various types.
Week 2	Python programming-I: Environments, toolkits incl. Jupyter, interfacing with other languages.
Week 3	Python programming-II: Rudimentary data structures, functions, file-processing for text and binary formats, and control flow. Locating data sources on the internet and data scraping using Python.
Week 4	Data wrangling and munging and data cleaning. First examples.
Week 5	Workflow Engineering - Types, construction, execution, etc.
Week 6	Python programming-III: Built-in data structures, functions, Numpy arrays, frames with pandas;
Week 7	Case Study I - String manipulation, workflows for unstructured textual and social media data using NLTK, bag-of-words, simple classifiers.
Week 8	Introduction to Javascript; Essential fundamentals and intro to D3.js.
Week 9	Word Clouds for text visualization; use of basic frequentist statistics and mapping visual attributes.
Week 10	Case Study II - Time Series analysis - moving averages, filtering, window-based data handling; visualization of univariate data;

Week 11	Case Study III - Tables and spatiotemporal maps; analysis and visualization;
Week 12	Case Study V - Analysis and display of image data;
Week 13	Case Study IV - Analysis and basic visualization of graphs;
Week 14	Course Recap and Prepare for Project Presentations

---

#### Academic Misconduct:

It is the responsibility of the Committee on Academic Misconduct to investigate or establish procedures for the investigation of all reported cases of student academic misconduct. The term “academic misconduct” includes all forms of student academic misconduct wherever committed; illustrated by, but not limited to, cases of plagiarism and dishonest practices in connection with examinations. Instructors shall report all instances of alleged academic misconduct to the committee (Faculty Rule 3335-5-487). For additional information, see the Code of Student Conduct <http://studentlife.osu.edu/csc/>.

---

#### Disability Services:

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. After registration, make arrangements with me as soon as possible to discuss your accommodations so that they may be implemented in a timely fashion. SLDS contact information: [slds@osu.edu](mailto:slds@osu.edu); 614-292-3307; [slds.osu.edu](http://slds.osu.edu); 098 Baker Hall, 113 W. 12<sup>th</sup> Avenue.

---

#### Title IX and Sexual Misconduct:

Title IX makes it clear that violence and harassment based on sex and gender are Civil Rights offenses subject to the same kinds of accountability and the same kinds of support applied to offenses against other protected categories (e.g., race). If you or someone you know has been sexually harassed or assaulted, you may find the appropriate resources at <http://titleix.osu.edu> or by contacting the Ohio State Title IX Coordinator, Kellie Brennan, at [titleix@osu.edu](mailto:titleix@osu.edu)

---

#### Diversity and Inclusion:

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 his or her 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.

### **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. The Ohio State University offers services to assist you with addressing these and other concerns you may be experiencing. If you or someone you know are suffering from any of the aforementioned conditions, you can learn more about the broad range of confidential mental health services available on campus via the Office of Student Life's Counseling and Consultation Service (CCS) by visiting [ccs.osu.edu](https://ccs.osu.edu) or calling [614-292-5766](tel:614-292-5766). CCS is located on the 4th Floor of the Younkin Success Center and 10th Floor of Lincoln Tower. You can reach an on call counselor when CCS is closed at [614-292-5766](tel:614-292-5766) and 24 hour emergency help is also available through the 24/7 National Suicide Prevention Hotline at 1-800-273-TALK or at [suicidepreventionlifeline.org](https://suicidepreventionlifeline.org).