Data 601: Introduction to Data Science

Spring 2019

General Information

## Meeting Times and Location

Wednesday, 7:10pm-9:40pm, at Information Technology [[TBD]]

|  |  |  |
| --- | --- | --- |
| Instructor | Email | Office Location & Hours |
| **Ben Payne** | benpayne@umbc.edu | ITE 201F, time [TBD](https://en.wikipedia.org/wiki/To_be_announced) |
| **I check my email daily and generally will respond to questions in the evening hours. If I do not respond to your email within 24 hours, please remind me that you are waiting for a response.**  **I am available before class for questions and help. Outside of that timeframe, please email me to schedule an appointment.** | | |

## Description

The goal of this class is to give students an introduction to and hands on experience with all phases of the data science process using real data and modern tools. Topics that will be covered include data formats, loading, and cleaning; data storage in relational and non-relational stores; data analysis using supervised and unsupervised learning using Python; data visualization; and scaling up for Big Data.

## Prerequisite

Students must be enrolled in the Data Science Program. Other students may be admitted with instructor permission. Students are expected to have experience with programming.

## Course Learning Objectives

Upon completion, students will understand:

* Understand issues relating to acquisition, cleaning and loading of data.
* Be able to perform data analysis using Python.
* Understand the basics of how data can be presented and visualized.
* Understand issues involved when the analysis scales up to Big Data.
* Understand pros and cons of the different data analysis methods
* Be able to provide estimates of requirements for storage and compute associated with analysis

# **Course Materials**

## Optional Texts

* "Data science from scratch: first principles with python" by Joel Grus. O'Reilly Media
* "Python Data Science Handbook" by Jake VanderPlas. O'Reilly Media
* "Data Wrangling with Python: Tips and Tools to Make Your Life Easier" by Jacqueline Kazil and Katharine Jarmul. O'Reilly Media
* "Think Like a Data Scientist: Tackle the data science process step-by-step" by Brian Godsey. Manning Publications

Please review options at the [UMBC library](https://library.umbc.edu/) . You do not need to buy any books. The PDFs available online.

## Recommended Software and Hardware

*All software used in this course is free*.

* Web browser capable of running [Jupyter Notebooks](https://en.wikipedia.org/wiki/Project_Jupyter).
* [Docker](https://en.wikipedia.org/wiki/Docker_(software)) for running containerized applications.
* [VirtualBox](https://en.wikipedia.org/wiki/VirtualBox) for running virtual computers.
* A laptop. Electrical outlets are available in the classroom. [UMBC Wi-Fi](https://doit.umbc.edu/services/network/) is available.
* Ability to connect your laptop to the classroom projector. [VGA](https://en.wikipedia.org/wiki/VGA_connector) and [HDMI](https://en.wikipedia.org/wiki/HDMI#Connectors) plugs are available. You can also use the classroom computer to display your content to the class.
* Paper and pen or pencil for in class exercises.

## Course Format and Assignments

Students will complete assigned homework, readings, essays, quizzes, two projects, and a final project. This course incorporates a variety of hands-on labs and practical exercises to engage students and prepare them for challenges they may encounter in the workplace.

Students will occasionally present their solutions to homework assignments in class. Projects will also involve presentations.

The final project will provide students opportunity to showcase what they have learned in a format similar to what they will encounter in a professional work setting.

## Course Communication

Email is preferred. I also use the Slack channel at <https://umbcdatasci.slack.com/messages> for addressing questions.

# Course Syllabus

Subject to revision; 20181215

|  |  |  |
| --- | --- | --- |
| Week 1 | Jan 30 | Overview of Data Science; Docker |
| Week 2 | Feb 6 | Python in Jupyter; Data structures |
| Week 3 | Feb 13 | Getting data; Ethics and Legality |
| Week 4 | Feb 20 | Data cleanup |
| Week 5 | Feb 27 | Project 1 presentations |
| Week 6 | March 6 | Automation; Reports |
| Week 7 | March 13 | Math (stats) |
| Week 8 | March 20 | No class – Spring Break |
| Week 9 | March 27 | Property graphs |
| Week 10 | April 3 | Elasticity, Cost/benefit analysis |
| Week 11 | April 10 | Scaling up |
| Week 12 | April 17 | Project 2 presentations |
| Week 13 | April 24 | TBD |
| Week 14 | May 1 | Clustering |
| Week 15 | May 8 | Regression |
| Week 16 | May XX | Final Project Presentations |

## Grading Criteria

Students are expected to participate in class discussions. Students will occasionally present results to the entire class.

|  |  |
| --- | --- |
| Course work | Grade Distribution |
| Attendance and Participation | 10% |
| Homework | 30% |
| Project 1 | 20% |
| Project 2 | 20% |
| Final Project and Presentation | 20% |

Final Grade will be computed as follows:

|  |  |
| --- | --- |
| 90-100% | A |
| 80 to 89% | B |
| 70 to 79% | C |
| 60 to 69% | D |
| <60 | F |

## Course Policies

UMBC provides a range of writing assistance, which can be found in the following:

* The Writing Center: <http://lrc.umbc.edu/tutor/writing-center/>
* Research Guides & Tutorials: <http://lib.guides.umbc.edu/tutorial>

Failure to follow guidelines for each assignment, including the required format, style, length, and submission may result in at least one-letter-grade reduction (10%) on the assignment depending on the type or number of transgressions.

Each assignment is provided with a required outcome, desired output format, and grading rubric. Read the rubric to earn the best score for your assignment. If you have questions, email the instructor.

Notify the instructor if you suspect you will not be able to complete the homework by the specified deadline. Late/incomplete assignments will be accepted if an extension has been agreed to in advance. Late submissions incur 10% reduction in score, with an additional 10% reduction per week. Emergency situations will be handled on a case by case basis with appropriate justification or documentation.

## Academic Integrity

By enrolling in this course, each student assumes the responsibilities of an active participant in UMBC’s scholarly community in which everyone’s academic work and behavior are held to the highest standards of honesty. Cheating, fabrication, plagiarism, and helping other to commit these acts are all forms of academic dishonesty, and they are wrong. Academic misconduct could result in disciplinary action that may include, but is not limited to failure, suspension or dismissal.

Refer to the UMBC policy on Academic Integrity: <http://catalog.umbc.edu/content.php?catoid=17&navoid=879#academic-integrity>.

Student Academic Misconduct in the Grad School is handled by the Associate Dean (AD) of the Grad School. The first step is for the instructor to consult with the AD. If the instructor and AD determine that a less serious infraction has occurred, then the AD provides written authority to the instructor to resolve the matter. The instructor then notifies the students in writing, and the students must be allowed the opportunity to provide an explanation. The student has the right to appeal the decision directly to the AD, and the student must be informed of this right. If an appeal is requested, then the AD will convene a Grievance Committee under the policy established by University of Maryland Graduate School, Baltimore (UMGSB).

Incidents are tracked by UMBC. Repeated infractions incur more severe penalties.

## Student Disability Services (SDS)

UMBC is committed to eliminating discriminatory obstacles that may disadvantage students based on disability. Services for students with disabilities are provided for all students qualified under the [Americans with Disabilities Act (ADA) of 1990](https://en.wikipedia.org/wiki/Americans_with_Disabilities_Act_of_1990), the [ADAAA of 2009](https://en.wikipedia.org/wiki/ADA_Amendments_Act_of_2008), and [Section 504 of the Rehabilitation Act](https://en.wikipedia.org/wiki/Section_504_of_the_Rehabilitation_Act) who request and are eligible for accommodations. The Office of Student Disability Services (SDS) is the UMBC department designated to coordinate reasonable accommodations that would allow students to have equal access and inclusion in all courses, programs, and activities of the University.

If you have a documented disability and need to request academic accommodations, please register with the Office of Student Disability Services (SDS) as soon as possible. To begin the registration process please visit the SDS website and review the registration information, including disability documentation guidelines and how to submit the SDS registration form online using the confidential data management software called Accommodate <https://sds.umbc.edu/accommodations/registering-with-sds/>.

Once accommodations have been approved, you and your instructors will be notified via an emailed accommodation letter from the SDS office. Both the SDS office and Shady Grove's [Center for Academic Success](https://shadygrove.umd.edu/student-services/center-for-academic-success)(CAS) will work with you to ensure you receive the approved accommodations. If you have any questions or concerns, please contact the [Office of Student Disability Services](https://sds.umbc.edu/) via [disAbility@umbc.edu](mailto:disAbility@umbc.edu) or phone at 410-455-2459. Please note that accommodations are not retroactive and begin once [SDS](https://sds.umbc.edu/) sends an approved accommodation letter.

For more information on the services CAS provides, please contact Mary Gallagher ([maryg@umd.edu](mailto:maryg@umd.edu)) or visit <https://shadygrove.umd.edu/student-services/center-for-academic-success>.

## Course Materials

All Course Materials created by the instructor are covered by a “[Creative Commons - Attribution 4.0 International](https://creativecommons.org/licenses/by/4.0/)” license.

Screen recordings of lectures, along with voice recording, will be used in class.

| Week | Reading | Exercises |
| --- | --- | --- |

| Homework |
| --- |

| Date | Subject |
| --- | --- |