Data 601: Introduction to Data Science

Fall 2018

General Information

## Meeting Times and Location

Thursday, 7:10pm-9:40pm, at Information Technology 229

|  |  |  |
| --- | --- | --- |
| 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. 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.

## 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/) and PDFs available online.

## Recommended Software and Hardware

* Web browser capable of running Jupyter Notebooks
* A laptop. Electrical outlets are available in the classroom. UMBC Wi-Fi is available.
* Paper and pen or pencil for in class exercises.

## Course Format and Assignments

The students will complete assigned homework, a midterm project and a final project with presentation. 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.

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

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

## Grading Criteria

Students are expected to participate in class discussions.

|  |  |
| --- | --- |
| Course work | Grade Distribution |
| Attendance and Participation | 10% |
| Homework | 30% |
| Midterm project | 30% |
| Final Project and Presentation | 30% |

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 on the assignment depending on the type or number of transgressions.

Late/incomplete assignments will be accepted if an extension has been agreed to in advance. Emergency situations will be handled on a case by case basis with appropriate justification or documentation.

Incomplete grades are granted only for extenuating circumstances and your request is made before the last week of class.

## 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 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 Syllabus

Subject to revision; 20180827

|  |  |  |
| --- | --- | --- |
| Week 1 | Aug 30 | Overview Data Science |
| Week 2 | Sept 6 | Python in Jupyter |
| Week 3 | Sept 13 | Math (stats) |
| Week 4 | Sept 20 | Regression |
| Week 5 | Sept 27 | Clustering |
| Week 6 | Oct 4 | Evaluation, cross-validation, overfitting |
| Week 7 | Oct 11 | *Substitute’s choice* |
| Week 8 | Oct 18 | Getting data |
| Week 9 | Oct 25 | Automation |
| Week 10 | Nov 1 | Data cleanup |
| Week 11 | Nov 8 | Scaling up |
| Week 12 | Nov 15 | Property graphs |
| Week 13 | Nov 22 | No class (Thanksgiving) |
| Week 14 | Nov 29 | Elasticity, Cost/benefit analysis |
| Week 15 | Dec 6 | Ethics and Legality |
| Week 16 | Dec 13 | Presentations |

| Week |  | Reading | Exercises |
| --- | --- | --- | --- |
| N/A |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

|  | Homework |  |  |
| --- | --- | --- | --- |
|  |  |  |  |
|  |  |  |  |
|  |  |  |  |

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