DATA 601: Introduction to Data Science, Fall 2022

Class Hours: M 6:30-9:10pm,
Class Room: Shady Grove Building IV 3305

Instructor: Felix Gonzalez, P.E.   Web: **Course's Github Page**
E-mail: fgonzale@umbc.edu   Office Hours: **Mondays, 6:00pm – 6:30pm**
Office: USG Adjunct Instructor Room, Building III, Room 4123

- If you want to secure a spot in my office hours, please email me 48 hours in advance.

- I respond to messages within 48 hours, excluding breaks and weekends.

- If your question is something about your code or something we covered in class you can schedule an appointment with our tutors. You can see our tutors' schedules at here.

Course Description

This class prepares students for more advanced topics in data science and introduces some of the tools and notions which are frequently used in the industry. Topics include: a review of Python programming and most fundamental modules; acquisition, handling, and working with different data; exploratory data analysis with statistics; data visualization and web scraping; life cycle of data science projects and different roles in a data science team, and ethical issues in data science.

Student Learning Outcomes

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

1. Describe the key activities in a data science project and understand the role of modeling in a data science project.

2. Use popular Python packages for exploratory data analyses, data visualization, and transformations.

3. Create functions and programs that clean, merge, and transform raw data sets and evaluate their quality for a given data science project.

4. Apply basic statistical knowledge in a data science project to test and verify hypotheses.

5. Apply best practices of communication for reporting upon completing a data science project.

Format and Procedures

This will be an in-person course and depending on the pandemic situation there might be online (synchronous) components. Computers with internet connection and working cameras are required for the online lectures. We will also be using a software called Jupyter Notebooks and/or Google

Collaboratory as the main medium of delivery for the lecture materials. For more information and to get familiarity with it, please check this tutorial. In addition to these:

1. Students will complete assigned homework, readings, quizzes.

2. Students will engage with hands-on labs and practical exercises to prepare them for challenges they may encounter in the workplace.

3. Students will occasionally present their solutions to homework assignments in class.

4. Students who are participating in the class online should be able to share both their video and audio.

Course Requirements

Textbook

We will not be following one single textbook in this course. Weekly reading materials and relevant course materials will be shared beforehand via Blackboard and/or the course's Github repository. In addition to these, I will be following the logical structures of the following textbook.

- Python Data Science Handbook: We will follow this book in the programming with Python parts of the course. The sections related to Pandas library are especially well written.

- Python for Data Analysis by Wes McKinney, O'Reilly, 2nd edition, 2017.

- Python Data Analytics With Pandas, NumPy, and Matplotlib, by Fabio Nelli, Apress, 2nd edition, 2018

Hardware Requirements

- Web browser capable of running Jupyter and/or Google Colaboratory Notebooks

- A computer with sufficient internet speed for online lectures. Make sure that your computer has video and microphone access.

Quizzes

There will be some quizzes in the beginning of the lectures to assess students' understanding of the reading assignments. Also, at the end of some lectures, students might be given quizzes to assess their understanding of the covered material.

Homework

There will be homework assigned to students roughly every two weeks. Depending on the scale of the homework, students will be given one or two weeks to submit their homework and other than exceptional circumstances this homework will be graded within a week. Please return your solution notebooks in Blackboard with the following filename convention: Lastname_HWXY.ipynb (e.g. Gonzalez_HW02.ipynb). If there is someone with the same last-name in the class, please add the initial of your first name between your last name and HWXY.

Attendance

In every lecture, I will take attendance at some point during the lecture. Students who miss this part of the lecture will be considered as absent and except for medical situations (or some other formal/written excuse) no excuse will be accepted for missing a class. Attendance will contribute to 5 percent of the final grade.

Grading

In the final grade, the assignments will have the following weights:

Attendance (14 Lectures)	—	5%
Quizzes (4 quizzes total)	—	9%
Homework (7 HW total)	—	56%
Projects (2 Projects total)	—	30%

Grading Distribution

Final letter grades will be assigned as follows: (Grades will be rounded upwards.)

94-100	—	A
93 - 90	—	A-
87 - 89	—	B+
83 - 86	—	B
80 - 82	—	B-
77 - 79	—	C+
73 - 76	—	C
70 - 72	—	C-
67 - 69	—	D+
0 - 59	—	F

Schedule and weekly learning goals

The schedule is tentative and subject to change.

Week 01, 09/05: Labor Day

Week 02, 09/12: Introduction to DATA601 and data science
- Course logistics, resources, and policies.
- Introduction and review of basic tools.
- Overview of the course: what you will learn and what we will not cover.
- Coding Basics: Naming conventions, objects and their methods, built-in functions etc.
- Introduction to Python and Jupyter Notebooks
- Jupyter Notebooks, Google Collaboratory, and Visual Studio Code
- HW: Install Anaconda Distribution or Test Run Google Collaboratory.

Week 03, 09/19: Python Programming
- Introduction to Variables in Python
- Conditionals
- Loops
- Input/output
- Functions
- Intro to Numpy

Week 04, 09/26: Pandas
- Dataframes, series, arrays, and lists
- Choosing an observation
- Choosing a variable
- Renaming, dropping a column, sorting values in a column
- Boolean mask and filtering
- Detecting missing values
- Describe, info, etc.

Week 05, 10/03: Basics of object-oriented programming
- Introduction to classes and objects
- Global variables
- Inheritance

Week 06, 10/10: Basics of visualization with Python:
- Visualization with matplotlib
- Figure, axis, titles, labels etc.
- Adding subplots to a figure
- Bar charts and box plots
- Advanced plotting

Week 07, 10/17: Data Transformations
- Operations on series: +/-/*, log, ranking, cutting etc.
- Grouping and aggregation
- Vectorized string operations
- Pivot tables
- Discuss Project 1

Week 08, 10/24: Exploratory data analysis
- Variation
- Typical values
- Unusual values
- Heatmaps
- Patterns and model

Week 09, 10/31: Loading and wrangling data
- Introductions to different data formats
- Reading csv, excel, html files
- Handling numbers
- Handling strings
- Tidying data

Week 10, 11/07: Working with date, time, and REX
- Time zone handling
- Moving Averages
- Resampling methods for time series
- Time series visualizations
- REX
- Reminder Project 1 due following week and Discuss Project 2

Week 11, 11/14: Relational data
- Merging, concatenation, joining datasets.
- Due date for Project 1, 11/14 23:59.

Week 12, 11/21: Introduction to Regression (1 of 3)
- Distributions
- Statistics with Python
- Introduction to Linear Regression

Week 13, 11/28: Introduction to Regression (2 of 2)
- Linear Regression (Cont.)
- Introduction to Logistic Regression
- Feature Selection

Week 15, 12/5: Version Control
- Git, Github
- Review Tools: Jupyter Notebook, Google Collaboratory and Visual Studio Code
- Due Date Project 2, 12/5, 23:59.

Week 16, 12/12: Review and Ethics
- Ethics, Privacy, and Data Science
- Class Review

Course Policies

During Class

Please be mindful that the COVID-19 pandemic is not over yet and everyone has different levels of concerns. This is why UMBC policies constitutes a common denominator that I expect everyone to follow in my classes. I ask everyone to be proactive about this and do their fair share in order to keep our community safe and protected. Moreover, if you have any concerns about this topic, please don't hesitate to share it with me immediately.

I understand that the electronic recording of notes will be important for class and computers will be allowed in class. Please refrain from using computers for anything but activities related to the class. Phones are prohibited as they are rarely useful for anything in the course. Eating and drinking are allowed in class but please refrain from it affecting the course. Try not to eat your lunch in class as the classes are typically active.

Policies on Incomplete Grades and Late Assignments

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.

Institutional Policies

Covid-19 Policies

Please see this Google doc for UMBC Policies and Resources during COVID-19.

Academic Integrity and Honesty

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: UMBC Academic Integrity Policy

Diversity Statement

It is my intent that students from all diverse backgrounds and perspectives be well-served by this course, that students' learning needs be addressed both in and out of class, and that the diversity that the students bring to this class be viewed as a resource, strength and benefit. It is my intent to present materials and activities that are respectful of diversity: gender identity, sexuality, disability, age, socioeconomic status, ethnicity, race, nationality, religion, and culture. Your suggestions are encouraged and appreciated. Please let me know ways to improve the effectiveness of the course for you personally, or for other students or student groups.

Student Disability Services

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, the ADAAA of 2009, and 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.

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(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 SDS via disAbility@umbc.edu or phone at 410-455-2459. Please note that accommodations are not retroactive and begin once SDS sends an approved accommodation letter. For more information on the services CAS provides, please contact Mary Gallagher (maryg@umd.edu) or visit Student-Services/Center for Academic Success

Title IX Statement

Any student who has experienced sexual harassment or assault, relationship violence, and/or stalking is encouraged to seek support and resources. There are a number of resources available to you. Please see The Office of Equity and Inclusion Website for recently updated UMBC Policies and Resources during COVID-19.

With that said, as an instructor, I am considered a Responsible Employee, as per UMBC's Interim Policy on Prohibited Sexual Misconduct, Interpersonal Violence, and Other Related Misconduct. This means that while I am here to listen and support you, I am required to report disclosures of sexual assault, domestic violence, relationship violence, stalking, and/or gender-based harassment to the University's Title IX Coordinator. The purpose of these requirements is for the University to inform you of options, supports, and resources.

You can utilize support and resources even if you do not want to take any further action. You will not be forced to file a police report, but please be aware, depending on the nature of the offense, the University may take action.

- The Counseling Center. phone: 410-455-27742 (M-F 8:30am - 5pm)

- University Health Services: 410-455-2542 (M-F 8:30am - 5pm) • For after-hours emergency consultation, call the police at 410-455-5555 Other on-campus supports and resources:

- The Women's Center(available to students of all genders): 410-455-2714 (M-Th 9:30am -6pm, F 9:30am - 4pm)

- Title IX Coordinator: 410-455-1606 (M-F 9am - 5pm)

If you need to speak with someone in confidence about an incident, UMBC has the following Confidential Resources available to support you: