DATA 601-06: Introduction to Data Science, Fall 2024

Class Hours: M 7:10-9:40pm,
Classroom: Shady Grove Building IV 3308

Instructor: Felix Gonzalez, P.E.
E-mail: fgonzale@umbc.edu

Web: **Course's Github Page**
Office Hours: **Mondays, 6:30 – 7:10pm**
Tuesdays, 12:15 – 1:00pm

Grader/Tutor: Sravya Chunduri
E-mail: mv54416@umbc.edu

Grader/Tutor Office Hours:
Wednesday, 5:00 – 6:30pm

Office: Classroom or if needed or specified by instructor, the USG Adjunct Instructor Room, USG, Building III, Room 4123

- If you want to secure a spot in outside of office hours, please email me 48 hours in advance.
- I try to respond to emails within 24 hours, excluding breaks and weekends.
- If your question is something about your code or something we covered in class, we have tutors available, you can request an appointment and tutors' schedules are available here.

Course Description

This class introduces topics to Python programming, data science concepts and prepares students for more advanced topics in data science. It provides an overview of the main 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; artificial intelligence, machine learning, and natural language processing; life cycle of data science projects, 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. Train and test fundamental machine learning algorithms.
6. 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 Collaboratory 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)	—	13%
Homework (7 HW total)	—	50%
Projects (2 Projects total)	—	32%

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

A detailed schedule of the weekly topics, learning goals, homework and projects can be found under the Class GitHub page in the file named "Data_Science_Introduction_Syllabus_Topics_Index.xlsx". The schedule is tentative and subject to change.

Week 01, 9/2: NO Classes, Labord Day Holiday

Week 02, 9/9:
- Class Logistics, requirements, and expectations
- Fundamental data science concepts and workflows
- Fundamental Python and Programming Concepts

Week 03, 9/16:
- Jupyter Notebook Development Environment Overview
- Introduction to Python and Built-in Functions

Week 04, 9/23:
- Python variables and data Structures (lists, tuples, set and dictionaries)
- Conditional statements and outputs
- While and For Loops
- Defining custom functions

Week 05, 9/30:
- Object oriented programming
- Jupyter Notebook Markdown Language
- Text and Regular Expressions (Regex)

Week 06, 10/7:
- Math with Numpy Library
- Introduction to Pandas

Week 07, 10/14:
- Working with Datetime
- Introduction to Data Analysis and Transformation with Pandas

Week 08, 10/21:
- Data Visualization with MatPlotLib
- Data Science Example: Story Telling and Data Cleaning

Week 09, 10/28:
- Data cleaning/wrangling/preparation for analysis
- Introduction to Exploratory Data Analysis (EDA)

Week 10, 11/4:
- Working with Files and Data Parsing
- Webservices, Application Programming Interface (API) and Web Scraping
- Relational Databases

Week 11, 11/11:
- Statistics: Fundamental of Statistics for Data Science
- Statistics: Hypothesis Testing

Week 12, 11/18:
- Introduction to Machine Learning and Regression
- Supervised ML Classification
- Classification Feature Selection

Week 13, 11/25:
- Introduction to Ethics in Data Science
- AI Example Demo Discussion Supervised and Unsupervised ML

Week 14, 12/2:
- Dashboarding
- Data Science Example: Dashboarding

Week 15, 12/9: LAST DAY OF CLASSES
- Project 2 Due
- Optional Lectures

Optional Lectures as time Permits:
- Introduction to Unsupervised ML: Clustering Algorithms
- Introduction to Natural Language Processing
- Other Special and Advanced Topics (e.g., Generative AI)

Student Engagement Resources at USG

The Universities at Shady Grove, in partnership with our Montgomery County educational partners Montgomery County Public Schools and Montgomery College, are committed to supporting students in their journey to become prepared for the world of work. We have collaborated with Montgomery County educators to develop an initiative called **Hire U** that contains a shared set of career readiness competencies which reflect the key behaviors that employers, across all industries and functional areas, are seeking in graduates entering the world of work.

These nine career competencies are adapted from best practices developed by the National Association of Colleges & Employers (NACE), in addition other sources, including employers. Employer involvement was key to the development of these national best-practice models, and including employer priorities is central to our approach in Montgomery County.

Complementing the skills and experiences students receive in the classroom, students must seek experiences out of the classroom, which will help to develop these competencies over time. The development of these competencies, in addition to completion of a college degree and participation in meaningful experiential learning experiences, will help students become prepared, competitive, and ready to thrive.

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

Use of AI Technical Assistance

Use of (i) any artificial technical assistance (i.e., ChatGPT or other generative technologies) outside of onboard spellcheck and things like Google Scholar, academic library databases, or reference managers and (ii) solutions/projects found on the internet are considered academic misconducts and strictly forbidden. If your solutions or reports are determined to be in this category, then you will receive 0 (zero) for that assignment/project.

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

Accessibility and Disability Accommodations, Guidance and Resources (required)

Accommodations for students with disabilities are provided for all students with a qualified disability under the Americans with Disabilities Act (ADA & ADAAA) 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 accommodations that creates equal access for students when barriers to participation exist in university courses, programs, or activities.

If you have a documented disability and need to request academic accommodations in your courses, please refer to the SDS website at sds.umbc.edu for registration information and office procedures.

SDS email: disAbility@umbc.edu

SDS phone: 410-455-2459

If you will be using SDS approved accommodations in this class, please contact the instructor to discuss implementation of the accommodations. During remote instruction requirements due to COVID, communication and flexibility will be essential for success.

Sexual Assault, Sexual Harassment, and Gender Based Violence and Discrimination (required)

UMBC Policy in addition to federal and state law (to include Title IX) prohibits discrimination and harassment on the basis of sex, sexual orientation, and gender identity in University programs and activities. Any student who is impacted by sexual harassment, sexual assault, domestic violence, dating violence, stalking, sexual exploitation, gender discrimination, pregnancy discrimination, gender-based harassment, or related retaliation should contact the University's Title IX Coordinator to make a report and/or access support and resources. The Title IX Coordinator can be reached at titleixcoordinator@umbc.edu or 410-455-1717.

You can access support and resources even if you do not want to take any further action. You will not be forced to file a formal complaint or police report. Please be aware that the University may take action on its own if essential to protect the safety of the community.

If you are interested in making a report, please use the Online Reporting/Referral Form. Please note that, if you report anonymously, the University's ability to respond will be limited.

Notice that Faculty and Teaching Assistants are Responsible Employees with Mandatory Reporting Obligations

All faculty members and teaching assistants are considered Responsible Employees, per UMBC's Policy on Sexual Misconduct, Sexual Harassment, and Gender Discrimination. Faculty and teaching assistants therefore required to report all known information regarding alleged conduct that may be a violation of the Policy to the Title IX Coordinator, even if a student discloses an experience that occurred before

attending UMBC and/or an incident that only involves people not affiliated with UMBC. Reports are required regardless of the amount of detail provided and even in instances where support has already been offered or received.

While faculty members want to encourage you to share information related to your life experiences through discussion and written work, students should understand that faculty are required to report past and present sexual harassment, sexual assault, domestic and dating violence, stalking, and gender discrimination that is shared with them to the Title IX Coordinator so that the University can inform students of their rights, resources, and support. While you are encouraged to do so, you are not obligated to respond to outreach conducted as a result of a report to the Title IX Coordinator.

If you need to speak with someone in confidence, who does not have an obligation to report to the Title IX Coordinator, UMBC has a number of Confidential Resources available to support you:

Retriever Integrated Health (Main Campus): 410-455-2472; Monday – Friday 8:30 a.m. – 5 p.m.; For After-Hours Support, Call 988.

Center for Counseling and Well-Being (Shady Grove Campus): 301-738-6273; Monday-Thursday 10:00a.m. – 7:00 p.m. and Friday 10:00 a.m. – 2:00 p.m. (virtual) Online Appointment Request Form

Pastoral Counseling via The Gathering Space for Spiritual Well-Being: 410-455-6795; i3b@umbc.edu; Monday – Friday 8:00 a.m. – 10:00 p.m.

Other Resources

Women's Center (open to students of all genders): 410-455-2714; womenscenter@umbc.edu; Monday – Thursday 9:30 a.m. – 5:00 p.m. and Friday 10:00 a.m. – 4 p.m.

Shady Grove Student Resources, Maryland Resources, National Resources.

Child Abuse and Neglect

Please note that Maryland law and UMBC policy require that faculty report all disclosures or suspicions of child abuse or neglect to the Department of Social Services and/or the police even if the person who experienced the abuse or neglect is now over 18.

Pregnant and Parenting Students

UMBC's Policy on Sexual Misconduct, Sexual Harassment and Gender Discrimination expressly prohibits all forms of discrimination and harassment on the basis of sex, including pregnancy. Resources for pregnant, parenting and breastfeeding students are available through the University's Office of Equity and Civil Rights. Pregnant and parenting students are encouraged to contact the Title IX Coordinator to discuss plans and ensure ongoing access to their academic program with respect to a leave of absence –

returning following leave, or any other accommodation that may be needed related to pregnancy, childbirth, adoption, breastfeeding, and/or the early months of parenting.

In addition, students who are pregnant and have an impairment related to their pregnancy that qualifies as disability under the ADA may be entitled to accommodations through the Office of Student Disability Services.

Religious Observances & Accommodations

UMBC Policy provides that students should not be penalized because of observances of their religious beliefs, and that students shall be given an opportunity, whenever feasible, to make up within a reasonable time any academic assignment that is missed due to individual participation in religious observances. It is the responsibility of the student to inform the instructor of any intended absences or requested modifications for religious observances in advance, and as early as possible. For questions or guidance regarding religious observances and accommodations, please contact the Office of Equity and Civil Rights at ecr@umbc.edu.

Hate, Bias, Discrimination and Harassment

UMBC values safety, cultural and ethnic diversity, social responsibility, lifelong learning, equity, and civic engagement.

Consistent with these principles, UMBC Policy prohibits discrimination and harassment in its educational programs and activities or with respect to employment terms and conditions based on race, creed, color, religion, sex, gender, pregnancy, ancestry, age, gender identity or expression, national origin, veterans status, marital status, sexual orientation, physical or mental disability, or genetic information.

Students (and faculty and staff) who experience discrimination, harassment, hate, or bias based upon a protected status or who have such matters reported to them should use the online reporting/referral form to report discrimination, hate, or bias incidents. You may report incidents that happen to you anonymously. Please note that, if you report anonymously, the University's ability to respond may be limited.

Other Class and Faculty References:

- Assistantships, Scholarships, and Financial Aid – Graduate Data Science Programs: Information Hub – UMBC
- Online Teaching – Planning Instructional Variety for Online Teaching – UMBC