DATS 4001 Data Science Capstone

Lecture 12 (48026) Spring 2026

Instructor information:

Name: Sarah Cassie Burnett
Email: sarah.burnett@gwu.edu
Lecture time: Wednesdays 3:30-6:00pm
Lecture location: Monroe Hall 250
Office hours: Thursdays 1:45-2:45pm or by appointment
Office location: Samson Hall 302

Course Description and Overview

Lecture, 2.5 hours; Capstone experience for data science majors. Application of theoretical knowledge and practical skills gained in major courses to a real-world problem. Review of ethical issues and current topics in data science. Restricted to data science majors.
Prerequisites: DATS 1001, DATS 2101, DATS 2102, DATS 2103, and DATS 2104.

This course is a capstone experience for data science majors that emphasizes the design, implementation, and communication of a computational data science product. Students will apply theoretical knowledge and practical skills developed throughout the major to build a reusable, well-structured tool, such as a Python package, modeling pipeline, or data analysis framework, designed to address a real-world data science problem.

The course emphasizes Python as the primary programming language, though students may use R or other appropriate tools where justified. Students will engage with the full data science workflow, including data acquisition, modeling, evaluation, and communication, while also developing strong software engineering practices such as version control, documentation, reproducibility, and code review.

Projects may be completed individually or in pairs, with clear expectations for division of labor and individual accountability. The course culminates in an in-class demonstration and a public-facing poster presentation showcasing each project.

Course Information

Students will create complex programs to solve problems of interest in data science and industry. Students will effectively deploy a range of packages to approach specialized tasks, and develop intuitive understandings of how these packages operate. Students will analyze and revise programs to improve clarity, robustness, and performance.

Upon completion of this course, students will be able to write, analyze, and communicate about Dats Science topics that accomplish a variety of complex computational tasks, including at least four of the following:

- Acquiring data sets via API, databases, or web-scrapers.
- Analyzing data using complex manipulation and visualizations.
- Researching the background of a particular field and summarizing how they quantify and/or classify data of interests with citations.
- Performing statistical inference and hypothesis testing (including methods such as linear regression and time series analysis) to support data-driven conclusions.
- Constructing complex machine learning pipelines for structured, image, and text data sets.
- Solving problems in computational mathematics, including systems of linear equations, eigenvalue problems and singular value decomposition, optimization, differential equations, and simulations.
- Deploying functional webpages using a Python-based (or R-based) web development framework.
- Constructing complex, robust, documented Python modules suitable for public dissemination and deployment.

Additionally, students will be able to identify and install Python (or R) modules to achieve a wider variety of tasks than those directly covered in the course.

Topics

- Advanced Python programming: multiple inheritance, decorators and wrappers, itertools, functools, multi-threading and multi-processing, setuptools.
- Open-Source Programming skills: version control with Git, understanding Python package structure, debugging, commenting, reading and writing documentation, code reproducibility
- Data Science skills and packages:
    - Review of EDA, inference, hypothesis testing
    - Web scraping
    - Web development with Flask
    - Deep learning with keras
    - SQL Database interaction
    - (maybe) Image processing

- (maybe) Graphic user interface with PyQt

Schedule (Tentative)

Week	Date	Topics	Assignment(s) Due
1	Jan 14	Capstone overview, Getting to each other Capstone GitHub Repo, rclone, LaTeX Ideas for a proposal	
2	Jan 21	Project Pitch share Intro to web, API	**Project Pitch** due January 20
3	Jan 28	Web scrapping w/ beautiful soup, regex, scrapy	**Project Proposal Final**
4	Feb 4	Weekly standup #1	Weekly update #1
5	Feb 11	Weekly standup #2	Weekly update #2
6	Feb 18		
7	Feb 25	Weekly standup #3	Weekly update #3
8	Mar 4	Weekly standup #4	Weekly update #4
9	Mar 11	**Spring Break (no class)**	
10	Mar 18	Weekly standup #5	Weekly update #5
11	Mar 25	Weekly standup #6	Weekly update #6
12	Apr 1	Weekly standup #7	Weekly update #7
13	Apr 8	Weekly standup #8 Data ethics review	Weekly update #8
14	Apr 15	Weekly standup #9	Weekly update #9
15	Apr 22	**Product Demonstrations**	Weekly update #10
	Apr 28	**Product Demonstrations (cont. if needed on this Makeup Day)**	
	May 1	**Public Poster Presentation, 11am – 3pm**	
	May 3	**Report due**	

Project proposal and weekly updates should be completed by Sunday 11:59pm that week. For example, the proposal will be due February 1.

Grading

Your final course grade will be calculated by:

Weight	Category
10%	Participation
15%	Weekly check-ins / Github updates
10%	Project proposal
15%	Product deliverable
10%	Product demonstration
20%	Poster and Poster presentation
15%	Final report
5%	Ethics and individual reflection

Grade Cut-offs

A [93, 100), A- [90, 93), B+ [87, 90), B [83, 87), B- [80, 83), C+ [77, 80), C [70, 77), etc.
Final cut-offs may be adjusted at the end of the quarter, but only to benefit students.

Generative Artificial Intelligence (GAI) in our Course

It might be the case that you want to use GAI to help you complete your product. It important that you use this to assist your understanding and not to do the work for you. I ask for two things when you do use it:

1. Cite it as you would for other references. Indicate where you used GAI for your code. Sample citation: "Create a script in Python to web scrape NBA.com season data" prompt. ChatGPT-4, OpenAI, 14 Mar. 2023, chat.openai.com/chat.
2. Thoroughly understand the code you use in your work. Rewrite/restructure it so that it makes sense to you. Seek help from yourself or others if you do not understand the code you are using. I will ask you what your codes mean and why.

Companies use AI but they also have code reviews to make sure that they have a quality product. As part of your product demonstration, I'll be asking a question related to product and code you've written.

University Policies

Observance of Religious Holidays

Students must notify faculty during the first week of the semester in which they are enrolled in the course, or as early as possible, but no later than three weeks prior to the absence, of their intention to be absent from class on their day(s) of religious observance. If the holiday falls within the first three weeks of class, the student must inform the faculty in the first week of the semester. For details and policy, see "Religious Holidays" at provost.gwu.edu/policies-procedures-and-guidelines.

Academic Integrity Code

Academic integrity is an essential part of the educational process, and all members of the GW community take these matters very seriously. As the instructor of record for this course, my role is to provide clear expectations and uphold them in all assessments. Violations of academic integrity occur when students fail to cite research sources properly, engage in unauthorized collaboration, falsify data, and otherwise violate the Code of Academic Integrity. If you have any questions about whether or not particular academic practices or resources are permitted, you should ask me for clarification. If you are reported for an academic integrity violation, you should contact the Office of Student Rights and Responsibilities (SRR) to learn more about your rights and options in the process. Consequences can range from failure of the assignment to expulsion from the university and may include a transcript notation. For more information, please refer to the SRR website (https://studentconduct.gwu.edu/academic-integrity), email rights@gwu.edu, or call 202-994-6757.

Use of Electronic Course Material and Class Recordings

Students are encouraged to use electronic course materials, including recorded class sessions, for private personal use in connection with their academic program of study. Electronic course materials and recorded class sessions should not be shared or used for non-course-related purposes unless express permission has been granted by the instructor. Students who impermissibly share any electronic course materials are subject to discipline under the Student Code of Conduct. Please contact the instructor if you have questions regarding what constitutes permissible or impermissible use of electronic course materials and/or recorded class sessions. Please contact Disability Support Services at disabilitysupport.gwu.edu if you have questions or need assistance in accessing electronic course materials.

Academic Support

Writing Center

GW's Writing Center cultivates confident writers in the University community by facilitating collaborative, critical, and inclusive conversations at all stages of the writing process. Working alongside peer mentors, writers develop strategies to write

independently in academic and public settings. Appointments can be booked online at gwu.mywconline.com.

Academic Commons

Academic Commons provides tutoring and other academic support resources to students in many courses. Students can schedule virtual one-on-one appointments or attend virtual drop-in sessions. Students may schedule an appointment, review the tutoring schedule, access other academic support resources, or obtain assistance at academiccommons.gwu.edu.

Support for Students Outside the Classroom

Disability Support Services (DSS) 202-994-8250

Any student who may need an accommodation based on the potential impact of a disability should contact Disability Support Services at disabilitysupport.gwu.edu to establish eligibility and to coordinate reasonable accommodations

Counseling and Psychological Services 202-994-5300

GW's Colonial Health Center offers counseling and psychological services, supporting mental health and personal development by collaborating directly with students to overcome challenges and difficulties that may interfere with academic, emotional, and personal success. healthcenter.gwu.edu/counseling-and-psychological-services.

Safety and Security

- In an emergency: call GWPD at 202-994-6111 or 911
- For situation-specific actions: review the Emergency Response Handbook at: safety.gwu.edu/emergency-response-handbook
- In an active violence situation: Get Out, Hide Out, or Take Out. See go.gwu.edu/shooterprep
- Stay informed: safety.gwu.edu/stay-informed