SYLLABUS

DSCI 611 ~ Data Workflow Automation ~ Winter 2026

Draft Syllabus

This is a draft copy of the syllabus that is not yet finalized. It is subject to change until publication to Canvas on Dec. 29, 2025.

AT A GLANCE

Course

Course Title	DSCI 611: Data Workflow Engineering
Credits	3 (min. 10 hrs/week of class and work)
Lecture (F2F)	Wednesdays 6–8:50 p.m., CCI 1056
Lecture (online)	Posted to Canvas on Thursdays
Readings	Posted to Canvas (from library & online articles)
Prerequisite	DSCI 511
Software	Git, Python, Data Version Control, additional packages as needed
Instructor	Michael Ekstrand mdekstrand@drexel.edu, CCI room 1113 Office hours by appointment (schedule here)

COURSE INFORMATION

Course Description

Teaches reproducible, adaptable automation for data processing, analytics, and modeling lifecycles for both scientific research and practical data science and predictive analytics in business settings. Covers data management and workflow automation tools, tracking and versioning data and analytics code, collaboration, integrating with data sources and machine learning deployments, inference servers, and experiment records. Course project applies the principles in an end-to-end data-intensive exercise.

Course Purpose Within a Program of Study

This course is a core component of the Big Data Analytics stackable certificate, and an elective for the MSDS program, that provides students with a foundation for creating high-quality, flexible, and reproducible data processing, analysis, and machine learning pipelines in their later courses and future work.

Course Learning Outcomes

Upon successful mastery of the course material, students will be able to:

1. Implement data processing, analytics, and modeling code in ways that can be robustly and automatically executed.
2. Apply design patterns such as ETL and ELT to create extensible, reproducible data processing workflows.
3. Design and develop multi-step automated workflows to reproducibly execute analyses and processes.
4. Store and manage input, intermediate, and final data with version control for sharing within a team or organization and between research and deployment environments.
5. Incorporate new data, models, and/or analyses into a workflow, evaluate proposed improvements, and monitor changes across the lifecycle of a data application.

COURSE MATERIALS

There is not yet a good textbook that systematically covers the core content of this course, so we will be assembling readings and resources from a range of sources. This includes the online documentation and tutorials for the software we are using, as well as additional online resources and books or book chapters.

Textbooks

There is no required text for this class. The following is a recommended textbook:

> Ben Wilson, *Machine Learning Engineering in Action*. Manning, 2022. ISBN: 978-1-61729-871-4.

Software Resources

We will use several current software packages throughout this course, and you will need to spend time with their documentation and tutorials. These include:

- Git
- Python
- Data Version Control (DVC)
- Pandas
- SciKit-Learn
- Jupyter
- Seaborn

LECTURE AND WEEKLY SCHEDULE

This course is dual-listed both in-person and online. Both sections will be run together as one Canvas course, primarily differentiated through lecture modality.

All Students

Pre-class readings, supplementary videos, and similar resources will be posted on Friday for the following week. I may add additional resources later in the week as well, based on class discussion and any final edits to my lecture material.

In-Person Lectures

In-person lectures will be **Wednesdays** from 6:00–8:50 p.m. in CCI 1056. If you are registered in-person, I strongly encourage you to attend all class sessions. If, however, you are unable to attend class due to illness or another pressing concern, please let me know, and you can watch the online lecture.

On Thursdays, I will post the lecture slides, examples, and other resources used in class. I will attempt to post example material in advance, but this will be on a best-effort basis.

Online Lectures

For online students, I will post each week's lecture on Thursday. Online students are also welcome to watch the lecture live on Wednesday nights through Echo 360. The recording will be lightly edited and will usually have better audio quality.

Discussion Boards

I encourage all students, both in-person and online, to use the Canvas discussion boards for asking questions about the class material or assignments, discussing class-related topics, and communicating with me and with each other.

ASSESSMENT PLAN

You will primarily demonstrate your learning through a **group project** with deliverables throughout the term. We will also have quizzes to check your knowledge and regular participation opportunities online and in-person. There is no midterm or final exam.

Table 1a shows how these components will contribute to your grade, and Table 1b details how final grades will be assigned.

Project

The main work of this course is a **project** that you will complete throughout the term, with deliverables every 1-2 weeks. This is a group project (in groups of up to 4, but I recommend 2-3 people), because collaboration is fundamental to the learning outcomes of the course (especially LO4). You will start the project by identifying data and a problem of interest to your group, and will progressively expand it out into a full data integration, modeling, and

(a) Course component grade contributions.

Category	Grade %
Project	70
Project Reflection	10
Quizzes	15
Participation	5

(b) Final grade scale.

Letter	-		+
A	90	93	98
B	80	83	88
C	70	73	78
D	60	63	68
F		0	

Table 1: Course grading details.

analysis pipeline by the end of the term. This project must be tracked in a private repository on either GitHub or Drexel CCI GitLab that is shared with me. The final project deliverables at the end of the term are:

- A **video presentation** shared with the class of your project and outcomes.
- The **final code** as a reproducible workflow in a Git repository.
- A **written report** and **project documentation** for your outcomes and how to run and extend your project.

All project deliverables will be submitted as a group. At the midpoint and end of the term, you will additionally submit an individual reflection on the learning and assessment of both you and your teammates' contributions to the project. The default will be for all team members to receive the same grade on the project deliverables, but I may adjust individual grades when there is evidence of significant imbalance in contribution (this is rare).

Since this is a dual-listed course, you are free to form a group comprising both in-person and online students.

Quizzes

There will be 5 short quizzes throughout the term to check your understanding of the material as we go, one every other week.

Participation

I expect you to engage with me and with your fellow students throughout the course. Each week, I will post discussion and reflection prompts for the discussion boards, and assign participation grades based on your active participation in those discussions (post at least 8 out of the 10 weeks of the term to earn full credit).

GETTING HELP

There is a good chance you will need help or clarification at some point during the semester. I provide several ways for you to ask questions and get support:

Discussion boards I recommend starting by posting a question to the Canvas discussion boards, as others may be confused about the same point you are, and then everyone can see the answer.

I also encourage you to answer other students' questions on the discussion boards. We often learn more by helping others, and I would rather see an excellent student answer that I can indicate is correct.

E-mail I am happy to answer questions and provide support by e-mail.

Office hours I do not have fixed office hours, but you can schedule time to meet with me either individually or with a group to discuss the course material, assignments, and your work. Use the link to my Bookings page in Canvas or on the first page of the syllabus to schedule a meeting with me. Please do not e-mail me to request a meeting, just schedule directly — if you e-mail, I will direct you to the bookings page, so you can save us both time. I can meet either in-person or over Microsoft Teams.

Availability

I try to respond to class questions within one business day. While I will sometimes work in evenings or weekends, I do not commit to responding to e-mails or discusison threads outside normal working hours (roughly 9–5 M–F).

COURSE POLICIES

Announcements

I will be using Canvas Announcements to communicate course announcements, including any changes to the syllabus or assignments. You are responsible for making sure that you receive course announcements in a timely fashion.

If I need to change the syllabus or an assignment description after its initial publication, I will include a dated Revision Log in that document describing the modifications.

Late Work

Each team has a budget of 4 **late days** to use at your discretion throughout the semester (with a max of 3 for any individual deliverable). Each day extends a project deliverable deadline by 24 hours, no questions asked. When using late days, mention in your assignment submission that you are using them. I appreciate advance notice that you plan to use them, but this is not required.

Late days are intended to cover most ordinary need for extensions, including such things as illness or child care needs, and additional individual extensions will rarely be granted.

Conduct

I expect you to respect me and your fellow students in all class interactions, both in official meetings such as lectures and out-of-classroom activities such as project group meetings and study sessions, and to contribute to a constructive learning environment.

In addition to the Drexel Conduct and Community Standards, the Recurse Center Social Rules are a good source of guidance on how to maintain a constructive and educational environment in a computing learning context.

Academic Integrity

As an information professional or data scientist, you are expected to do your own work, attribute sources, and respect the legal and moral rights of others with respect to their work; as a student, you are also required to abide by the university policies regarding academic integrity. While I aim to allow you to make reasonable use of resources, cheating (including copying code, using unauthorized resources during tests, etc.) is not acceptable. If I find you to be cheating, the penalty may range from an F on the assignment to an F on the course, and will be reported to the university.

You may consult external resources such as other books and web sites for understanding how to solve assignments or portions of the project. In your assignment submissions, list all external resources you used; if they are available online, provide the URL. You do not need to cite the textbooks or the official documentation for the software we are using. You must specifically declare any use of generative AI (e.g. ChatGPT or GitHub Copilot); see the Artificial Intelligence policy below for further details on use of AI.

Besides the course forum, you may ask questions related to the course material and concepts required to complete the work on archived, publicly-accessible discussion forums such as Stack Overflow, newsgroups, or publicly-archived mailing lists. To qualify as publicly-accessible, a site must provide access to complete discussions without requiring payment or registration. Provide URLs to the forum discussion on the relevant web site or archive with your assignment submission. When you ask a question for one of the assignments, mention that it is for a course project and that your instructor permits you to make reasonable use of discussion forums.

A good question will ask about how to go about a particular sub-portion of the problem, how something works, why something you are trying doesn't work, or other specific difficulties. Do not ask "how do I solve <the problem description>?", or similarly direct translations of the project requirements, or for specific code. Questions should be written to fill in a gap in your understanding that will then enable you to continue your work, not

to get a solution to the assignment. The same principles apply if you are attempting to use an AI chatbot to get help with understanding the material or complete the assignments.

This course is also subject to the Drexel Academic Integrity, Plagiarism, Dishonesty, and Cheating Policy; the above does not replace or modify that policy, but is intended to clarify how I apply the university policy to this course.

Permitted Use of Artificial Intelligence

You are allowed to use generative AI tools such as ChatGPT and GitHub Copilot in the assignments and project for this course, although I do not encourage you to do so. If you use generative AI, you must do the following:

- State that you used generative AI, including which model(s) you used.
- Clearly document which portions of your submitted work were produced, drafted, or substantially edited by AI.
- Save a record of your prompts and system responses, and provide it as an appendix to your submission.

The Drexel Policy on Academic Integrity Pertaining to Artificial Intelligence provides further details on university policies regarding AI.

AI and Citations

I require you to *declare* your use of AI, but not to *cite* AI as a source. AI itself is not a citable source of information, as it is not an original source of information (either primary or secondary; it is at best tertiary), and generative AI tools rarely provide stable, referencable identifiers for a reader to refer to the exact infromation you received.

Disability Accommodations

If you need particular accommodations to be able to fully participate in this course, please talk with me as soon as possible. If you have documentation from Disability Resources for particular accommodations, please bring it, but I am happy to discuss with you anything needed for you to fully participate in the class.

Office Hours

My office hours are by appointment, arranged through my Microsoft Bookings page. There is a link to this page at the beginning of the syllabus and in Canvas.

Course Changes

I may need to make changes to the course as the term progresses to better support your learning and the logistics of delivering the course. Such changes will be announced through Canvas Announcements, as well as in lecture when timing permits.

SCHEDULE

The following is the *tentative* schedule of topics, along with due dates for the term. I may need to adjust the topic schedule as we progress through the term; Canvas will be the current, authoritative reference. Due dates are fixed unless I make a specific announcement.

Week	Lecture	Topic	Quiz	*Deliverable*
1	1/7	Introducing Reproducibility and Automation		
2	1/14	Version Control and Collaboration	1	*Team + Topic*
3	1/21	Wrangling and Processing Data		*Initial Setup*
4	1/28	Robust, Configurable, Automatable Code	2	
5	2/4	Training Models		*EDA*
6	2/11	Evaluation, Reporting, and Monitoring	3	
7	2/18	Tuning and Experimentation		*First Models*
8	2/25	Parallel Processing and Acceleration	4	
9	3/4	Deployment and Production		*Experiments*
10	3/11	Documenting, Reporting, and Publication	5	*Presentation*
F	3/18	*Finals Week*		*Final Report*

Deliverables are due **Sunday at midnight** at the end of the week in which they are listed.