DATA 400: Capstone in Data Analytics

Spring 2026

(Last updated January 23, 2026; syllabus is subject to change)

Instructor: Professor Eren Bilen
Office: Althouse 104
Email: bilene@dickinson.edu
Phone: 717-254-8162
Office Hours: Calendly link
            Monday 11:30a-1p
            Wednesday 11:30a-1p
            or by appointment
QRA: Alaina Rongione, rongiona@dickinson.edu
     Minh Trinh, trinhm@dickinson.edu
QRA Office Hours:
Alaina, TBA
Minh, TBA

Classroom: Althouse G7
Tuesday and Friday
S1: 1:30-2:45p,
S2: 3-4:15p

Required Course Materials

- Ace the Data Science Interview: 201 Real Interview Questions Asked By FAANG, Tech Startups, and Wall Street; by Kevin Huo and Nick Singh
- BIG Jobs Guide: Business, Industry, and Government Careers for Mathematical Scientists, Statisticians, and Operations Researchers; by Rachel Levy, Richard Laugesen, Fadil Santosa
- The Effect, by Nick Huntington-Klein: Link
- Course webpage: Github

Course Overview

This capstone course aims to provide students a foundation for the must-have skills needed towards becoming a successful and ethical data analyst. The course must be completed successfully as the final core requirement for degree in data analytics. During the course, we will revisit topics that you have learned in your prior data analytics core courses. **By the end of the course, you will produce a clear output:** a data science project showcasing your data analysis skills. You should consider this course and the project as an opportunity to practice your oral presentation, writing, and time management skills. Additionally, we will discuss strategies that will help you navigate the job market in data analytics.

A student who successfully completes this course should be able to do the following:

- Develop research questions systematically that are feasible to implement in a reasonable timeline, evaluate what makes a research question "good" or "bad",

- Evaluate and characterize the limitations and uncertainty of statistical inference and machine learning models,
- Develop the ability to write about and present results from a data analysis project to a general and/or technical audience including its ethical implications.

Course Policies

Attendance Policy: This course will be taught in person. Students are expected to attend all in-class meetings. While I will not take formal attendance, it is important for you to attend in class meetings and take notes. If you will be unable to attend a class meeting for any health-related issues or other emergencies, please contact me beforehand so that arrangements can be made.

Use of Laptops, Tablets, and Phones: Laptops and tablets are permitted for note-taking and coding during this course. In exchange for trusting you to use these devices, I ask that you not use them as distractions. I maintain the right to change this policy for individual students or for everyone if these tools become a problem during class. Phones are not permitted and should be put away in silent mode

Grading: Your course grade is based on assignments that are tied closely with a semester long course project. First, you will submit two potential research questions that will involve data analysis. These questions can be ideas that you came up with during your previous courses, or can be new ones. A grading rubric to be used to evaluate your ideas will be provided beforehand. You will receive feedback from me and your peers on your ideas.

The first idea that you submit will form the basis of a "mini project" which you will work on individually during the first four weeks of the course. **Your mini-project must be related to your three course disciplinary focus.** In Week 5, you will present your findings from your mini-project, which should include exploratory data analysis and viz, optionally a model. Next, you will move onto your "big project", which can be a continuation of your mini-project, or could be derived from another idea you submitted. This project will be more involved, and must include both exploratory data analysis and an appropriate machine learning model.

Your final product will include a ReadMe guide explaining your data + how you obtained it, and a final presentation + slides. Your final presentation will be graded on tech skills (e.g., your model and how you implemented it) and soft skills (organization and clarity of your slides, delivery of your talk).

Your participation in in-class discussions is important. You will be assigned readings (see the last page of the syllabus), and will be graded on how well you prepare and contribute to in-class discussions. The weights of the assignments are:

Research questions (one-page) (x2) (15%): Due dates on Pg.6
Mini-project presentation (10 mins) (15%): Due dates on Pg.6
Progress presentation (10 mins) (15%): Due dates on Pg.6
Data collection + ReadMe guide (15%): Due dates on Pg.6
Poster session (15%): Due dates on Pg.6
Final presentation (15 mins) (15%): Due dates on Pg.6
Participation in discussions (10%)

The following scale will be used to determine your final grade:

Score	Letter	GPA	Score	Letter	GPA
$93 \geq x$	A	4.0	$73 \leq x < 77$	C	2.0
$90 \leq x < 93$	A-	3.7	$70 \leq x < 73$	C-	1.7
$87 \leq x < 90$	B+	3.3	$67 \leq x < 70$	D+	1.3
$83 \leq x < 87$	B	3.0	$63 \leq x < 67$	D	1.0
$80 \leq x < 83$	B-	2.7	$60 \leq x < 63$	D-	0.7
$77 \leq x < 80$	C+	2.3	$x < 60$	F	0.0

Teams: It is important that you gain practice in team-work, as data analysts often work in teams. You will be asked to form <u>teams of two</u> by the end of Week 6. A spreadsheet where you can enter your teams (along with your team names) will be provided. Between Week 3 and Week 6, you should chat with your classmates and find someone you share interests with. So, be on the lookout for what others are working on for their mini-project, get together, discuss what you can do for your "big project" as a team.

Datacamp: You will receive an invitation to Datacamp.com, which will grant you access to all their tutorials on data science for the duration of the semester. Use this as an opportunity to brush up the skills that might be necessary for your project, or to gain practice before your interviews.

Writing in the Discipline (WiD): This course is part of the WiD thread in the Data Analytics Major. Data scientists regularly report their analyses to their teams and other stakeholders. The types of writing that data scientists do takes on many different forms for a variety of audiences: graphical/visual representations, briefs, memos, reports, academic papers, code, data documentation, with audiences varying from developers, internal and external technical audiences to general audiences among others. Students completing the Data Analytics Major satisfy the WiD graduation requirement by maintaining a portfolio of assignments from the courses on the WiD Thread. This course will include one output (your data ReadMe guide) which will be added to your WiD portfolio, focusing specifically on:

- Communicating clearly the motivation and the goals of a data science project, providing a clear strategy for how the data was obtained,

- Clearly and transparently documenting and annotating data sources, data wrangling, and variable descriptions.

Moreover, an unsatisfactory WiD portfolio scores an automatic F in this course. The data analytics major at Dickinson achieves the Writing in the Discipline (WiD) graduation requirement via a collection of targeted assignments in several required courses.

Each student stores these assignments in a WiD portfolio, and they officially achieve the WiD graduation requirement when that portfolio is assessed to be satisfactory by the senior seminar instructor. The instructor will make an initial determination of whether a student's WiD portfolio is satisfactory by the end of the second week of classes. If it is unsatisfactory, the student will be given a chance to rectify the portfolio by completing one or more additional relevant WiD assignments. If the WiD portfolio is still unsatisfactory at the end of the semester, the student will receive an automatic F in this course.

Ethics: Data scientists must be aware of the ethical implications of their work. It must be the case that a data science project or algorithm objectively improves the well-being of the society without introducing bias, unfairness, or a violation of privacy. We will critically evaluate each of the ideas you propose and your project from such ethical lens, and discuss case studies that will help you become more aware of the ethical issues that you will most likely face working as a data analyst.

Make-up: If you must be away from campus on university business or due to an emergency, or you will miss a presentation day due to sickness, please reach out to me as soon as you can and provide documentation. The weight of the missed assignment will be shifted to other assignments.

Getting Help

Office Hours: Please use this `Calendly link` to schedule an office meeting so that you have a confirmed slot scheduled with me. I will be holding office hours each week. Please see the first of page of the syllabus for my hours. I am also available by appointment. If there is a conflict and you are unable to make it to any of my hours, please feel free to send me an email. My availability outside office hours is not guaranteed, however I devote my attention fully to you during my office hours. Therefore, I highly encourage you to come to my office hours and ask questions.

Other Important Information

Generative Artificial Intelligence (Gen-AI) Statement: Use of Gen-AI (such as Chat-GPT) is allowed in this class, but I encourage you to use it responsibly. Research shows that Gen-AI can help you learn more quickly, *if* you are already engaged with the material. It also shows that students with a solid coding foundation benefit more from using Gen-AI, so it is important for you to invest in your coding skills. My recommendation is that if you are going to use Gen-AI, treat it as a tutor, not a solver. While you are learning new concepts, it can be helpful for explanations or clarifications, but you should avoid using it to complete full assignments for you. Most importantly for our class, while Gen-AI can serve as a useful editor, it is not good at generating original ideas. Its suggestions tend to reflect average or typical content from its training data. For creative or analytical work, your ideas should come from you, not from a statistical average.

Referencing the Work of Others: When submitting your work, you must follow common-sense ground rules. You must cite all your resources in any delivery of your work. Failure to cite your references constitutes plagiarism. If you have any questions about what constitutes plagiarism, you should reach out to me, or you may contact the writing center. Here is a nice resource that you can refer to for any questions regarding plagiarism.

Statement on Disabilities: Dickinson values diverse types of learners and is committed to ensuring that each student is afforded equitable access to participate in all learning experiences. If you have (or think you may have) a learning difference or a disability – including a mental health, medical, or physical impairment – that would hinder your access to learning or demonstrating knowledge in this class, please contact Access and Disability Services (ADS). They will confidentially explain the accommodation request process and the type of documentation that Dean and Director Marni Jones will need to determine your eligibility for reasonable accommodations. To learn more about available supports, go to www.dickinson.edu/ADS, email access@dickinson.edu, call (717) 245-1734, or go to the ADS office in Room 005 of Old West, Lower Level (aka "the OWLL").

If you have already been granted accommodations at Dickinson, please follow the guidance at www.dickinson.edu/AccessPlan for disclosing the accommodations for which you are eligible and scheduling a meeting with me as soon as possible so that we can discuss your accommodations and finalize your Access Plan. If test proctoring will be needed from ADS, remember that we will need to complete your Access Plan in time to give them at least one week's advance notice.

SOAR: Academic Success Support: Students can find a wealth of strategic guidance by going to www.dickinson.edu/SOAR. This website for SOAR (Strategies, Organization, and Achievement Resources) includes apps, tips, and other resources related to time management, study skills, memory strategies, note-taking, test-taking, and more. You will also find information aimed to helps students "SOAR Through Academic Challenges," as well as a schedule of academic success workshops offered through Academic Advising. If you would like to request one-on-one assistance with developing a strategy for a manageable and academically successful semester, email SOAR@dickinson.edu.

Course Outline: Below is a schedule for what will be covered in this course and the deadlines. There may be adjustments on the list during the semester depending on progress made in class. On Tuesdays, we will have an in-class discussion/mini-lecture on the week's topic. Thursdays will consist of a project meeting session during which you will work on your project in class and ask me any questions that you might have about your project. Due dates for each assignment will be announced.

Select tech topics: You will cast your vote for what topic we will discuss during Week 10. Possible topics we can cover: Web-scraping, Artificial Neural Networks (ANNs), Geospatial data with Geopandas, Text analysis/Natural Language Processing (NLP), Advanced Regression, Causality & Natural Experiments. Feel free to email me if you have more suggestions.

Other talks, events: I highly encourage you to actively attend campus talks organized by the departments of Data Analytics, Math & CS, Economics, among others. You should network with presenters, and chat about how they approached tackling on their project. Your attendance on these talks will also count towards your participation. We will have a discussion on what each talk was about and how the presenter collected and analyzed their data in the following class.

Guest visits: I will coordinate and invite Dickinson alumni to our class so you can learn from their experiences and network with them. I highly encourage you to come and ask

any questions you might have since it is a great opportunity to learn from people who are active in the field. Your contribution in discussions during these guest visits will also count towards your participation. So far we have two guest speakers scheduled. See Page 6 for the dates of their visits.

Course outline

Date	Topic	Readings	+ to-do	Due*
Week 1	Data Analyst job postings, Github portfolio	ADSI Ch 1,2- 1, 2	work on job market items	
Week 2	How to ask good research questions?	1, 2	work on idea 1	idea 1
Week 3	Revisiting Data 180 + case studies	1, 2	start mini-project + work on idea 2	idea 2
Week 4	Revisiting Data 200 + case studies	ADSI Ch 8	cont. mini-project	
Week 5: **Feb 17 & Feb 20**	**Mini-project presentations**		come + provide feedback	slides by Feb 17 @1:30p
Week 6	Revisiting Phil 258 + case studies	1, 2, 3, 4, 5	form teams + start big-project	teams
Feb 24 @1:30p, Feb 27 @3p	**Alum visit TBA: Bio**		come + ask questions	
Week 7	Tech topic 1	notes on github	cont. big-project	
Week 8	*Spring break*			
Week 9	Revisiting Data 300 + case studies	ADSI Ch 7, 1	cont. big-project	
Week 10: **March 24 & 27**	**Progress presentations**		come + provide feedback	data sub (final.)
Week 11	Tech topic 2	notes on github	cont. big-project	
Week 12	Poster prep session		cont. big-project	
April 8, by 5p	**Poster abstract submission**			
April 11, by 5p	**Poster submission** `link`			
Week 13	Interactive visualization: Bokeh, Streamlit	1, 2, 3, 4	cont. big-project	data readme
April 16, Th @4:30p	**Poster day**		come + provide feedback	
Week 14	Data Science: The video game		cont. big-project	slides due
Week 15	Wrap up projects		cont. big-project	
S1: May 5, M @2p S2: May 8, T @2p	**Final presentation**		come + provide feedback	

ADSI: Ace the Data Science Interview

∗: each item is due on Friday of the week at 5p, unless indicated otherwise.