

DATA 400: Senior Seminar in Data Analytics

Spring 2023

(Last updated August 11, 2022; syllabus is subject to change)

Instructor: Professor Eren Bilen

Classroom: TBA

Office: Rector North 1309

Email: bilene@dickinson.edu

Phone: 717-254-8162

Office Hours: TBA

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 Lauge-sen, Fadil Santosa
- The Effect, by Nick Huntington-Klein: [Link](#)
- Course webpage: [Github](#)

Course Overview

This capstone course aims to provide students a strong 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 core data analytics 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 for 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
- Distinguish correlations from causality, outline methodology and concisely explain results, and characterize the limitations and uncertainty of statistical inference and machine learning algorithms
- Develop the ability to write about and express an opinion on an ethical issue in data analytics for a broad general and/or technical audience

Course Policies

Attendance Policy: This course will be taught in person in (TBA). Students are expected to attend all in-class meetings, which occur on (TBA). 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 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 three potential research questions involving 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 peer feedback and my feedback for each of your ideas. You will then select one idea that you would like to pursue further and provide a more detailed presentation to the class in form of a proposal presentation. Next, you will provide a technical presentation, related to specific tools you have to use/have used in order to complete your project (more info on this later).

Your final product will include a ReadMe guide explaining your data and how you obtained your data and a final presentation. Your final presentation will be graded on the organization and clarity of your slides, and the delivery of your talk. You will have an opportunity to present your project to other students on campus during the student research symposium day. You will be assigned readings that we will discuss in class from which you will be graded on how well you prepare and your contribution to the discussions. The weights of the assignments are:

Research questions (one-page) (x3) (15%):	Due dates TBA
Proposal presentation (10-15 mins) (15%):	Due dates TBA
Progress report (1-2 pages) (5%):	Due dates TBA
Progress presentation (5-10 mins) (10%):	Due dates TBA
Data collection + ReadMe guide (15%):	Due dates TBA
Poster session (15%):	Due dates TBA
Final presentation (10-15 mins) (15%):	Due dates TBA
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: You will be asked to form teams of two during the first two weeks of the course. You may pursue your project individually, if you wish to do so.

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 assignment (your final presentation) that will be added to the 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 will be obtained.
- Clearly and transparently documenting and annotating data sources, data wrangling, and variable descriptions.

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: I will be holding office hours in-office. Please see the course page for the Zoom link that will be used throughout the semester. My office hours are (TBA). I am also available by appointment.

Quantitative Reasoning Center

Dickinson College provides additional support for students taking courses with quantitative content across the curriculum through the Quantitative Reasoning (QR) Center. For the Spring 2023 semester, the QR Center will offer tutoring for our course, in addition

to general quantitative support. You are strongly encouraged to make an appointment with them. [Click here](#) to access the QR Center webpage.

Please visit dickinson.mywconline.com to make an appointment. Then, access the drop-down menu under “limit to” at the top of the scheduler and select DATA 400. This will restrict the tutor list and schedule to only those tutors approved for this course. When you make your appointment, please also paste or upload your assignment and any work that you have done.

Other Important Information

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 help 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 Mondays, we will have an in-class mini-lecture based on the week’s topic. Thursdays will consist of project meeting sessions during which you can work on your project in class and ask me any questions you might have about your project.

Date	Topic	Readings	Due
Week 1	Data Analyst job postings, Github portfolio	ADSI Ch 1,2- 1, 2	
Week 2	How to ask good research questions?	1, 2	teams
Week 3	Revisiting Data 180	1, 2	idea 1
Week 4	Revisiting Data 200	ADSI Ch 8	idea 2 & 3
Week 5	Revisiting Phil 258	1, 2, 3	data sub (recomm.)
February 23	Proposal presentations		
Week 6	Revisiting Data 300	ADSI Ch 7	
Week 7	Select tech topic 1	NLP, ANN, Geodata	
March 9	Progress report		
Week 8	<i>Spring break</i>		
Week 9	Select tech topic 2	NLP, ANN, Geodata	data sub (final)
Week 10	Interactive visualization: Tableau, Shiny	1, 2, 3, 4	
Week 11	Interactive visualization: Bokeh, Streamlit	1, 2, 3, 4	
April 6	Progress presentations		
Week 12	Data Science: The video game	Link TBA	
Week 13	Wrap up projects		data readme
Week 14	Wrap up projects		
April 27, Thursday @5pm	Poster day		
Week 15	Wrap up projects		
May 9, Tuesday @2pm	Final presentation		

ADSI: Ace the Data Science Interview

Important Dates for the Spring 2023 Semester

Last Day to Add/Drop or Change to/from Pass/Fail	Sunday, January 29
Spring Vacation	Monday, March 13 thru, Friday, March 17
Last Day to Withdraw from a Course with a "W" grade	Friday, April 21
Classes End	Friday, Friday, May 5
Reading Period Days	May 6, 7