

# DATA 451: Data Science Capstone

## Winter 2019

### Course Syllabus

January 6, 2019

<b>Instructor:</b>	Alexander Dekhtyar	Hunter Glanz
<b>email:</b>	dekhtyar@calpoly.edu	hglanz@calpoly.edu
<b>office:</b>	14-210	25-111

What	Day	Time	Location
Meeting Times	MWF	10:10 am – 12:00pm	14-257

#### Office Hours

	Alex		Hunter	
	When	Where	When	Where
Monday	9:10am - 10:00am	14-210		
Tuesday	1:10pm - 3:00pm	14-210		
Wednesday	9:10am - 10:00am	14-210	12:10pm-2:00pm	25-111
Friday			12:10pm-2:00pm	25-111

Additional appointments can be scheduled by emailing the instructors at [dekhtyar@calpoly.edu](mailto:dekhtyar@calpoly.edu) and or [hglanz@calpoly.edu](mailto:hglanz@calpoly.edu)

## Course Sections

This course is team-taught by Hunter Glanz and Alex Dekhtyar. For logistical reasons, there are two sections of the course, one with Hunter as the instructor of record and one with Alex. Some of you are registered for Alex's section, some — for Hunter's.

**All of you are in the SAME course.**

This means that **it makes no difference** to the instructors, which section you are in, and **it will make no difference to you**. Only **one course** is taught. **All requirements, expectations, assignments, grading, etc...** are **the same** for all students taking this course.

If you need to discuss course material with the instructors, you can contact either of the us, regardless of who is listed as the instructor of record for your

section (in fact, the decisions on who to talk to probably will be based on the specific complementary expertise of the instructors).

## Overview

This is the first quarter of a two-quarter Data Science capstone course. During this course you will

- Get assigned to a project team
- Work with the project customer to outline the project requirements
- Design a variety of solutions appropriate to the problem(s) your team will be studying
- Begin implementing your solutions
- Write a progress report and give a mid-project presentation in class
- Work on a number of assignments designed to improve your writing
- Receive some additional instruction on machine learning techniques not covered in DATA 401.

We have six hours of contact time/week. This time is broken down as follows:

Team meetings:	2.5 hours
Meeting with instructors/client:	0.5 hours
Lecture:	1 hour
Writing assignments/discussion:	1 hour
Miscellaneous activities:	1 hour

Starting Week 2, each team will have a specific time slot for meetings with customer/instructors (even weeks - customer meeting, odd weeks - meeting with instructors w/o customers).

## Textbook and Literature

The official textbook for the course is

Joseph M. Williams and Joseph Bizup. *Style: Lessons in Clarity and Grace*. (Any edition is fine.) Pearson, 2016.

Please **go ahead and buy this book on-line** as soon as possible, as the first assignment will come at the end of week 1 of the class.

## Topics and Tentative Schedule

Table 1 contains the tentative schedule for the course. Please note, that oral presentations will take place during Week 10 of classes. There are no in-person meetings/presentations scheduled for the finals week, *however, your written reports will be due some time during the finals week*.

## Grading

Your overall grade will be calculated according to the following weighting scheme.

<b>Weekly Progress Reports</b>	<b>20%</b>
<b>End-Quarter Presentation</b>	<b>20%</b>
<b>End-Quarter Report</b>	<b>20%</b>
<b>Peer Assessment</b>	<b>20%</b>
<b>Assignments</b>	<b>20%</b>

Week	Activity
1	Introduction, Project Presentations
2	Team formation, initial meetings
3 - 5	Requirements, Data Collection, Feature Engineering
6 - 9	Initial analysis
10	Report and presentation preparation, Mid-project presentations
Finals week	Report due

Table 1: Tentative schedule for the course.

## Course Participation expectations

We expect everyone to actively participate in the work on the team projects. Your level of contribution to your teams will be reflected in your course participation grade as well. To this end, we will conduct post-project peer assessments, where contributions of each person will be peer-reviewed by their teammates.

## Communication

Because there are two sections, Dr. Glanz does not have the ability to e-mail students in Dr. Dekhtyar's section, nor does Dr. Dekhtyar have the ability to e-mail students in Dr. Glanz's. To make communication easier, we will post all announcements on the Piazza forum.

If sending questions to the instructors, we encourage you to address the questions to both of us. One of the instructors may have a better insight than the other into a specific issue you might have, although at times, collective wisdom of both instructors will yield a much better answer than individual responses would have. Additionally, both instructors would want to know what inquiries are made, and how they were responded to.

## Web Page

Class web page can be found at

<http://www.csc.calpoly.edu/~dekhtyar/DATA451-Winter2019>

Through this page you will be able to access all class handouts including homeworks, project information and lecture notes. Despite the URL, the web page will be maintained by both instructors.

## Academic Integrity

### University Policies

Cal Poly's Academic Integrity policies are found at

<http://www.academicprograms.calpoly.edu/academicpolicies/Cheating.htm>

In particular, these policies define *cheating* as (684.1)

*"...obtaining or attempting to obtain, or aiding another to obtain credit for work, or any improvement in evaluation of performance, by any dishonest or deceptive means. Cheating includes, but is not limited to: lying; copying from another's test or examination; discussion of answers or questions on an examination or test, unless such discussion is specifically authorized by the instructor; taking or receiving copies of an exam without the permission of the instructor; using or displaying notes, "cheat sheets," or other information devices inappropriate to the prescribed test conditions; allowing someone other than the officially enrolled student to represent same."*

Plagiarism, per University policies is defined as (684.3)

“... the act of using the ideas or work of another person or persons as if they were one’s own without giving proper credit to the source. Such an act is not plagiarism if it is ascertained that the ideas were arrived through independent reasoning or logic or where the thought or idea is common knowledge. Acknowledgement of an original author or source must be made through appropriate references; i.e., quotation marks, footnotes, or commentary.”

University policies state (684.2): “Cheating requires an “F” course grade and further attendance in the course is prohibited.” (appeal process is also outlined, see the web site above for details.). Plagiarism, per university policies (684.4) can be treated as a form of cheating, although a level of discretion is given to the instructor, allowing the instructor to determine the causes of plagiarism and effect other means of remedy. It is the obligation of the instructor to inform the student that a penalty is being assessed in such cases.

### Course Policies

**Work.** Individual work is to be completed by individual students. Team projects are to be completed by the combined team. No people outside the class shall be involved in completing any assignments. Some level of consultation between students in the class on team projects (and on some of the individual assignments) is permitted, but all work claimed on project reports and in other deliverables must be completed by the team itself, and any substantive contributions of classmates from other teams<sup>1</sup> must be explicitly acknowledged in all writeups in this class.

**Code sharing and Licensing Compliance.** Use of third-party libraries for data analysis, visualizations, and other tasks (data cleaning, e.g.) is generally permitted, subject to the appropriateness of the licensing terms to the particulars of the project<sup>2</sup>. Similarly, your data acquisition efforts must conform to any explicit rules and limitations set forth by the data owner(s).

In all cases where you are not sure how to proceed, **PLEASE CONSULT THE INSTRUCTORS!**

**Citations.** In all written work, please follow the standards of proper academic citations. All submitted papers must feature full bibliography containing all references cited in the both of the submission. Direct quotes from academic and non-academic sources *must be represented explicitly as quotes* in your submissions, and must be properly attributed. Failure to do so will be considered an instance of plagiarism.

---

<sup>1</sup>For example, if a member of another team helped you brainstorm the architecture of a neural network, or if your team used a data parsing tool built by another team and shared with yours.

<sup>2</sup>E.g., GPL-licensed tools cannot be used for a product that your customer may want for commercial use, although they can be used for non-commercial prototypes.