

Carleton University
Department of Systems and Computer Engineering
SYSC4906 Introduction to Data Science – Winter 2021
Course Outline

Instructor Information and Office hours

Dr. Hoda Khalil

Office Hours: By appointment on Wednesdays 11:35 - 12:55

e-mail: hodakhalil@cunet.carleton.ca

TA Information and Office hours

To be determined. See CULearn for more details.

Calendar Description

At the discretion of the Department, a course dealing with selected advanced topics of interest to students in Biomedical and Electrical, Communications, Computer Systems, Electrical, Software Engineering, and Engineering Physics may be offered.

Prerequisite(s): permission of the Department.

<http://calendar.carleton.ca/undergrad/courses/SYSC/>

Prerequisites

Permission of the Department.

SYSC 2006 and (ECOR 2050, SYSC 2510, or STAT 3502)

Assumed Knowledge

Students should

- Understand basic statistical concepts.
- Have mathematical skills in linear algebra and probabilities.
- Have knowledge of python or be prepared to acquire the basics of python independently.

Course Objectives

This course is not intended to cover all principles of data science in detail. Instead, it provides a wide overview of the main concepts in data science for beginners. It introduces a set of preliminary tools and techniques to perform data science tasks.

By the end of the course, students will learn the basics of the different properties of data (structure, size, and type) and will be able to categorize data based on their properties. Students will be able

to choose and perform data munging operations for a variety of purposes. They will also gain an introductory level understanding of analytical techniques that will allow them to infer insights from data. Students will learn data visualization methodologies and will be able to choose and evaluate data visualizations that are appropriate for answering a given research question. Besides, students will learn how to apply some software engineering principles and concepts such as anticipation of change and separation of concerns in the context of data science. Students are expected to prioritize ethical values of data science when performing any of the acquired skills and techniques.

Learning Outcomes

By the end of the course students will be able to:

1. formulate the data science question and collect the data relevant to it.
2. demonstrate their understanding of the different properties of data (structure, size, and type) and will be able to categorize data based on their properties.
3. choose and perform data munging operations for a variety of purposes.
4. use common data science tools for reading, filtering, manipulating, and plotting data.
5. apply descriptive statistical analysis within the context of a data science project.
6. demonstrate an introductory level understanding of analytical techniques that will allow them to infer insights from data.
7. apply data visualization methodologies, choose, and evaluate data visualizations that are appropriate for answering a given research question.
8. design, create, validate, and evaluate simple data models.
9. apply some software engineering principles and concepts such as anticipation of change and separation of concerns in the context of data science.
10. prioritize ethical values of data science when performing the acquired skills and techniques.

Graduate Attributes (GA's)

The Canadian Engineering Accreditation Board requires graduates of engineering programs to possess 12 attributes at the time of graduation. Activities related to the learning outcomes listed above are measured throughout the course and are part of the department's continual improvement process. Graduate attribute measurements will not be taken into consideration in determining a student's grade in the course. For more information, please visit:

<https://engineerscanada.ca/>.

Please include all GA indicators mapped to your course and the learning outcomes relation to these GA's in a table similar to the sample provided below.

Graduate Attribute	Learning Outcome (s)
<p>SAMPLE ONLY 1.8.S Knowledge Base: Discipline-Specific Concept SCE-5: <i>Software Engineering</i> 2.1 Problem Analysis: <i>Problem Definition</i></p>	<p>1-7 8</p>

Textbooks (or other resources)

The primary text that will be used for the course is:

Skiena, Steven S. *The Data Science Design Manual*. 1st ed. 2017. Springer International Publishing, 2017.

Secondary Book:

Igual, Laura., and Santi. Seguí. *Introduction to Data Science A Python Approach to Concepts, Techniques and Applications*. 1st ed. 2017. Springer International Publishing, 2017.

Evaluation and Grading Scheme

Participation in discussion forums: 10% (5 contributions)

Quizzes: 10% (5 quizzes)

Assignments: 30% (planning phase: 10%, delivery phase: 15%, and presentation 5%)

Midterm: 20%

Final: 30%

Breakdown of course requirements

Due to COVID-19, availability during class time is not mandatory, except for a few course elements that are detailed below. The course material will be delivered using a blended model involving synchronous and asynchronous components.

Lectures: Lectures will be prerecorded and posted on CULearn by the assigned class time (asynchronous), except for the first lecture in the course and the two lectures that involve guest speakers which will be delivered synchronously. Although synchronous material will also be recorded and posted on CULearn, after the class assigned time, attendance is highly encouraged.

Discussion Forums: There will be two types of discussion forums on CULearn course page: (1) Discussion Forum A: The instructor will post questions that are related to the course material 5 times during the whole term. You are required to answer 3 out of the 5 posted set of questions. (2) Discussion Forum B: In this forum, you can ask questions about any lecture topic. You are required to participate at least twice in this forum either by posting a question or by commenting on or answering a question posted by one of your peers. This sums up to at least 5 posts during the term.

Quizzes: Five (5) quizzes will be delivered using CULearn quiz tool and will be administered during class time. If any student has a problem attending the quiz during class time, they should contact the instructor or the TA.

Project/Assignments: You are required to complete one project in two layered milestones. The first milestone is the project planning phase, while the second is the project delivery phase and a presentation. The project is to be completed in groups of 5 students. The TA will facilitate 2 optional meetings throughout the term, during the lecture time, to help students meet and to provide support for their projects' tasks. The meeting time will be considered part of the lecture. More details concerning the project will be posted on CULearn.

Midterm: The midterm will be during class time using CULearn quiz tool. If you are unable to attend the midterm during the assigned time for a valid reason, the weight of the midterm mark will be added to the final exam.

Final Exam: The final exam will be scheduled during the formal final examination period. The final exam is for evaluation purposes only and will not be returned to students. You will be able to make arrangements with the instructor or with the department office to see your marked final examination after the final grades have been made available.

Tentative Week-by-Week breakdown

Week	Date	Topic	Instructor's deliverable	Students' deliverable
1	11-1-2021	What is data science?	Course outline	Read course outline
	13-1-2021	Data properties	Discussion Forum A 1	
2	18-1-2021	Data scientist toolbox		
	20-1-2021	Formulating questions		Discussion Forum A 1
3	25-1-2021	Collecting data		
	27-1-2021	Data preparation I+ Quiz 1	Discussion Forum A 2	Quiz 1
4	1-2-2021	Data preparation II + project meeting		
	3-2-2021	Descriptive statistics I		Discussion Forum A 2
5	8-2-2021	Descriptive statistics II		
	10-2-2021	Visualizing data + Quiz 2		Quiz 2
	12-2-2021			Project planning phase
6	15-2-2021	Winter Break		
	17-2-2021			
7	22-2-2021	Inferential statistics I	Discussion Forum A 3	
	24-2-2021	Inferential statistics II + Quiz 3		Quiz 3
8	1-3-2021	Guest Speaker I		Discussion Forum A 3
	3-3-2021	Midterm		
9	8-3-2021	Basics of Machine Learning		
	10-3-2021	Regression Analysis + Quiz 4	Discussion Forum A 4	Quiz 4
10	15-3-2021	Reporting results		
	17-3-2021	Network analysis I		Discussion Forum A 4
11	22-3-2021	Network analysis II		
	24-3-2021	Big data I + project meeting	Discussion Forum A 5	
12	29-3-2021	Big data II + Quiz 5		Quiz 5
	31-3-2021	Data Ethics		Discussion Forum A 5
13	5-4-2021	Guest Speaker II		
	7-4-2021	Project presentations		Project delivery phase
14	12-4-2021	Project presentations		
	14-4-2021	Project presentations		

General Regulations

For the few recorded lectures, you will be notified at the start of the session when the recording will start. Please note that recordings are protected by copyright. The recordings are for your own educational use, but you are not permitted to share or publish any course material including recordings to third party sites, such as social media sites and course materials sites.

Attendance: Students are expected to attend all lectures and lab periods. The University requires students to have a conflict-free timetable. For more information, see the current *Undergraduate Calendar, Academic Regulations of the University, Section 2.1.3, Course Selection and Registration and Section 2.1.7, Deregistration*. **Due to COVID-19, availability during class time is not mandatory, except for a few course elements that are detailed above.**

Health and Safety: Every student should have a copy of our Health and Safety Manual. A PDF copy of this manual is available online: <http://sce.carleton.ca/courses/health-and-safety.pdf>

Deferred Term Work : Students who claim illness, injury or other extraordinary circumstances beyond their control as a reason for missed term work are held responsible for immediately informing the instructor concerned and for making alternate arrangements with the instructor and in all cases this must occur no later than three (3.0) working days after the term work was due. The alternate arrangement must be made before the last day of classes in the term as published in the academic schedule. For more information, see the current *Undergraduate Calendar, Academic Regulations of the University, Section 4.4, Deferred Term Work*.

Appeal of Grades : The processes for dealing with questions or concerns regarding grades assigned during the term and final grades is described in the *Undergraduate Calendar, Academic Regulations of the University, Section 3.3.4, Informal Appeal of Grade and Section 3.3.5 Formal Appeal of Grade*.

Academic Integrity: Students should be aware of their obligations with regards to academic integrity. Please review the information about academic integrity at: <https://carleton.ca/registrar/academic-integrity/>. This site also contains a link to the complete Academic Integrity Policy that was approved by the University's Senate.

Plagiarism: Plagiarism (copying and handing in for credit someone else's work) is a serious instructional offense that will not be tolerated.

Academic Accommodation: You may need special arrangements to meet your academic obligations during the term. You can visit the Equity Services website to view the policies and to obtain more detailed information on academic accommodation at <http://www.carleton.ca/equity/> For an accommodation request, the processes are as follows:

- **Pregnancy or Religious obligation:** Please contact your instructor with any requests for academic accommodation during the first two weeks of class, or as soon as possible after the need for accommodation is known to exist. For more details

see <https://carleton.ca/equity/wp-content/uploads/Student-Guide-to-Academic-Accommodation.pdf>

- **Academic Accommodations for Students with Disabilities:** The Paul Menton Centre for Students with Disabilities (PMC) provides services to students with Learning Disabilities (LD), psychiatric/mental health disabilities, Attention Deficit Hyperactivity Disorder (ADHD), Autism Spectrum Disorders (ASD), chronic medical conditions, and impairments in mobility, hearing, and vision. If you have a disability requiring academic accommodations in this course, please contact PMC at 613-520-6608 or pmc@carleton.ca for a formal evaluation. If you are already registered with the PMC, contact your PMC coordinator to send me your *Letter of Accommodation* at the beginning of the term, and no later than two weeks before the first in-class scheduled test or exam requiring accommodation (*if applicable*). **Requests made within two weeks will be reviewed on a case-by-case basis.** After requesting accommodation from PMC, meet with me to ensure accommodation arrangements are made. Please consult the PMC website (www.carleton.ca/pmc) for the deadline to request accommodations for the formally-scheduled exam (*if applicable*).
- **Survivors of Sexual Violence:** As a community, Carleton University is committed to maintaining a positive learning, working and living environment where sexual violence will not be tolerated, and where survivors are supported through academic accommodations as per Carleton's Sexual Violence Policy. For more information about the services available at the university and to obtain information about sexual violence and/or support, visit: <https://carleton.ca/sexual-violence-support/>.
- **Accommodation for Student Activities:** Carleton University recognizes the substantial benefits, both to the individual student and for the university, that result from a student participating in activities beyond the classroom experience. Reasonable accommodation must be provided to students who compete or perform at the national or international level. Please contact your instructor with any requests for academic accommodation during the first two weeks of class, or as soon as possible after the need for accommodation is known to exist. For more details, see <https://carleton.ca/senate/wp-content/uploads/Accommodation-for-Student-Activities-1.pdf>

Copyright on Course Materials: The materials created for this course (including the course outline and any slides, posted notes, labs, project, assignments, quizzes, exams and solutions) are intended for personal use and may not be reproduced or redistributed or posted on any web site without prior written permission from the author(s).