

**School of Information Management
INFO 6270 Introduction to Data Science
Winter 2020/2021**

Course Type (e.g. F2F, online, blended): Online and asynchronous, with synchronous options (see Instructional Methods)

Instructor: Colin Conrad

Office: Rowe 4020

Telephone: (902) 494-8378

E-mail: Colin.Conrad@dal.ca

Preferred method of contact: Microsoft Teams for all technical questions, Email for other questions

Office hours: By appointment only, using email or MS Bookings (link [here](#))

Course website: Brightspace

Tutorials: To be determined (by popular choice)

Teaching Assistant name/contact info: Pallavi Gone

ABOUT YOUR INSTRUCTOR



Colin Conrad, Ph.D. | Assistant Professor
School of Information Management
Dalhousie University

Website: <https://colinconrad.com>

How to get in touch with me: I check and respond to emails at least once daily between 9 am and 10:30 am Atlantic time (though often more frequently) every weekday. I try hard to respond all emails within 48 hours. In addition, I will regularly check on MS Teams to answer questions and participate in the discussion, though will try to limit my activity on Teams to certain days of the week. If you would like to schedule a one-on-one meeting, please reserve a time using the MS Bookings app or email me to schedule a time.

COURSE DESCRIPTION

This course serves as an introduction to data science, an increasingly important set of skills and techniques for business intelligence, effective governance, and the research process. The amount of data we generate increases year by year. As computers have begun to play roles in many aspects of our daily life, our actions and interactions leave digital traces. This has led both to an explosion in the amount of

data that we generate and an increased interest in analyzing and understanding that data. This class will give you an introduction to the skills you need to effectively collect, manipulate, and analyze data yourself.

Rather than being constrained to using any specific data analysis software, we will focus on using the flexible programming language Python. You will receive a thorough introduction to Python, learning how to use a variety of its built-in capabilities as well as a number of available data analysis packages. By the end of this course, you should be capable enough to begin teaching yourself and expanding your data science skills.

COURSE PRE-REQUISITES

None, though the onus is on you to identify whether this course is the right fit. This course is designed for a keen information management (or other management-type) students with little technical background but interest in learning new technical skills.

I strongly recommend that you have at least taken a course in information systems (e.g. INFO 5590) before taking this course. It would also be an asset if you have also taken a course with a database component (e.g. INFO 6540), though this course does not directly build on that content. You are not expected to know how to program before taking this course, though *must* be prepared to learn complex technical content. In fact, this course is not appropriate for students whose undergrad degree is in computer science or for students who are already proficient in programming. The course is conducted almost entirely using computer programming but structured in a way to equip students who have little-to-no programming experience with the skills necessary to do basic data science.

LEARNING OBJECTIVES

This course will introduce students to computer programming, computational thinking, data analysis and basic statistics programming. The instruction will help students become more familiar with using their computer to conduct computational tasks—especially data-oriented tasks. By doing so, it will give students a foundation from which to build upon in order to learn more advanced and specialized computational research and data analytic skills in the future.

LEARNING OUTCOMES

Upon completion of the course, students are expected to have gained basic knowledge or proficiency in the following areas:

1. Understand the principles of data analysis;
2. Learn the basics of data manipulation (sometimes referred to as data wrangling or data munging);
3. Achieve a foundation in data visualization skills;
4. Become competent with the Python programming language;
5. Become familiar with SQL;
6. Achieve a critical appreciation for computational methods and computational thinking;
7. Learn about the IT requirements for data analysis;

8. Plan and implement an information management framework;
9. Understand how to construct and maintain datasets.

TECHNOLOGY USED

Virtually all technical components of this course will be conducted using Jupyter notebooks running the Python programming language running through packages provided by a package called Anaconda. This software is entirely open source and can be downloaded [using this link](#). A detailed list of the technologies are given below:

- Unit 1: Introduction to Python for Data Science (Python, Jupyter)
- Unit 2: Data Science Basics (Python, Jupyter, Pandas, NumPy, SciPy, Scikit-Learn)
- Unit 3: Other Topics in Data Science (Python, Jupyter, SQLite, Tweepy, R)

Again, all of these technologies are provided in most recent Anaconda distributions so students will not need to worry about installation of technologies beyond Anaconda.

INSTRUCTIONAL METHODS

All class lectures are provided online asynchronously and can be accessed through your Brightspace account and are released at 8:00 am on each Monday at the beginning of the week. All course deliverables will be due on 11:55 pm on a Friday. You will also have the option to attend a live help session held at a popular time determined early in the semester.

In Dalhousie's MI program, students come from many diverse backgrounds and may not have deep experience with technology. You are not expected to be a technical expert and need not be an expert to perform well in this course. The goal of this course is to introduce students to the skills required to be effectively use and manage information technologies that are used in most organizations (whether they be non-profit, government, or private). Students who do well in this course nonetheless consistently demonstrate an openness to synthesising hands-on experience with managerial considerations that are relevant to their interests or career context.

LEARNING MATERIALS

This course does not have a textbook, though we will draw from one textbook for the first module. This textbook can be downloaded for free from the author's website or purchased in print online:

Sweigart, A. (2020). *Automate the boring stuff with python: Practical programming for total beginners*. Retrieved from: <https://automatetheboringstuff.com/>

Every week we will use electronic learning materials developed in Jupyter provided by the instructor. There will be readings provided throughout the course which could be useful to completing your course deliverables.

METHODS OF EVALUATION

Detailed instructions regarding each assignment will be provided. Assessment of all assignments is directly related to attention to the instructions, clarity of expression and presentation, and evidence of significant analysis and reflection.

See also the [SIM Grading Policy](#).

COMPONENT	DETAILS	DUE DATE	VALUE
Weekly In-Class Exercises	Every week, we will work on small in-class exercises together following the presentation and scrum meeting. 10 of the 12 exercises must be submitted on Brightspace for 5 points each. The exercises are due 11 days following their release, always on a Friday at 11:55 pm.	Every following Friday, after class	50%
Individual Project	Propose and complete a small independent project using the skills developed in this course. You have the option of uploading your scripts in the SIM SUCA (“Shut Up and Code Afternoon”) repository for the benefit of future students and Dalhousie as a whole, if you so choose.	Apr 12 th	50%

INTEGRATION OF [MI COMPETENCIES](#)

PROGRAM COMPETENCY	COURSE LEARNING OUTCOME	COURSE ASSESSMENT
Information Management Leadership	1, 6, 8	IP
User-centred Information Services	1, 2, 3, 4, 5, 6, 7	WCE, IP
Management of Information Technology	1, 2, 3, 4, 5, 6, 7, 8, 9	WCE, IP
Research and Evaluation	2, 3, 4, 6, 9	WCE, IP
Risk Management	7, 8, 9	IP
Change Management	7, 8, 9	IP
Workplace Skills & Attributes:	1, 2, 3, 4, 5, 6, 7, 8, 9	WCE, IP

CLASS POLICIES

Attendance

Class attendance is required in all MI courses and is included in the weekly class exercises mark.

Citation Style

SIM courses use APA as the default standard citation style. Unless the instructor provides alternative written instructions, please use the APA citation style in your assignments to briefly identify (cite) other people's ideas and information and to indicate the sources of these citations in the References list at the end of the assignment. For more information on APA style, consult Dalhousie Library website at <https://libraries.dal.ca/help/style-guides.html> or the APA's Frequently Asked Questions about APA

Late penalties for assignments

A penalty for late assignments will be assessed, unless prior permission has been given by the instructor to submit an assignment late, which normally will be for extended illness, medical, or family emergencies only (see below). Late submissions will be assessed a penalty of five percent per day, including weekends. Assignments will not normally be accepted seven days or more after the due date; in such cases the student will receive a grade of zero.

Missed or Late Academic Requirements due to Student Absence:

Dalhousie University recognizes that students may experience short-term physical or mental health conditions, or other extenuating circumstances that may affect their ability to attend required classes, tests, exams or submit other coursework.

Dalhousie students are asked to take responsibility for their own short-term absences (3 days or less) by contacting their instructor by phone or email prior to the academic requirement deadline or scheduled time **AND** by submitting a completed [Student Declaration of Absence form](#) to their instructor in case of missed or late academic requirements. Only 2 separate Student Declaration of Absence forms may be submitted per course during a term (note: faculty, college, school, instructor or course-specific guidelines may set a lower maximum).

SIM GRADING POLICY

A+	90-100	Demonstrates original work of distinction.
A	85-89	Demonstrates high-level command of the subject matter and an ability for critical analysis.
A-	80-84	Demonstrates above-average command of the subject matter.
B+	77-79	Demonstrates average command of the subject matter.
B	73-76	Demonstrates acceptable command of the subject matter.
B-	70-72	Demonstrates minimally acceptable command of the subject matter.
F	<70	Unacceptable for credit towards a Master's degree.

ACCOMMODATION POLICY FOR STUDENTS

Students may request accommodation as a result of barriers experienced related to disability, religious obligation, or any characteristic protected under Canadian human rights legislation.

Students who require academic accommodation for either classroom participation or the writing of tests and exams should make their request to the Advising and Access Services Center (AASC) prior to or at the outset of the regular academic year. Please visit www.dal.ca/access for more information and to obtain the Request for Accommodation form.

A note taker may be required as part of a student's accommodation. There is an honorarium of \$75/course/term (with some exceptions). If you are interested, please contact AASC at 494-2836 for more information or send an email to notetaking@dal.ca.

Please note that your classroom may contain specialized accessible furniture and equipment. It is important that these items remain in the classroom, untouched, so that students who require their usage will be able to fully participate in the class.

ACADEMIC INTEGRITY

In general:

The commitment of the Faculty of Management is to graduate future leaders of business, government and civil society who manage with integrity and get things done. This is non-negotiable in our community and it starts with your first class at Dalhousie University. So when you submit any work for evaluation in this course or any other, please ensure that you are familiar with your obligations under the Faculty of Management's Academic Integrity Policies and that you understand where to go for help and advice in living up to our standards. You should be familiar with the [Faculty of Management Professor and Student Contract on Academic Integrity](#), and it is your responsibility to ask questions if there is anything you do not understand.

Dalhousie offers many ways to learn about academic writing and presentations so that all members of the University community may acknowledge the intellectual property of others. Knowing how to find, evaluate, select, synthesize and cite information for use in assignments is called being "information literate." Information literacy is taught by Dalhousie University Librarians in classes and through Dalhousie Libraries' online [Citing & Writing](#) tutorials.

Do not plagiarize any materials for this course. For further guidance on what constitutes plagiarism, how to avoid it, and proper methods for attributing sources, please consult the University Secretariat's [Academic Integrity](#) page.

Please note that Dalhousie subscribes to plagiarism detection software that checks for originality in submitted papers. Any paper submitted by a student at Dalhousie University may be checked for originality to confirm that the student has not plagiarized from other sources. Plagiarism is considered a very serious academic offence that may lead to loss of credit, suspension or expulsion from the University, or even the revocation of a degree. It is essential that there be correct attribution of authorities from which facts and opinions have been derived. At Dalhousie, there are University Regulations which deal with plagiarism and, prior to submitting any paper in a course; students should read the Policy on [Academic Dishonesty](#) contained in the Calendar.

Furthermore, the University's Senate has affirmed the right of any instructor to require that student assignments be submitted in both written and computer readable format, e.g.: a text file or as an email attachment, and to submit any paper to a check such as that performed by the plagiarism detection software. As a student in this class, you are to keep an electronic copy of any paper you submit, and the course instructor may require you to submit that electronic copy on demand. Use of third-party originality checking software does not preclude instructor use of alternate means to identify lapses in

originality and attribution. The result of such assessment may be used as evidence in any disciplinary action taken by the Senate.

Finally:

If you suspect cheating by colleagues or lapses in standards by a professor, you may use the confidential email: ManagementIntegrity@dal.ca which is read only by the Assistant Academic Integrity Officer.

Faculty of Management clarification on plagiarism versus collaboration:

There are many forms of plagiarism, for instance, copying on exams and assignments. There is a clear line between group work on assignments when explicitly authorised by the professor and copying solutions from others. It is permissible to work on assignments with your friends but only when the professor gives you permission in the specific context of the assignment. University rules clearly stipulate that all assignments should be undertaken individually unless specifically authorised.

Specific examples of plagiarism include, but are not limited to, the following:

- Copying a computer file from another student, and using it as a template for your own solution
- Copying text written by another student
- Submitting the work of someone else, including that of a tutor as your own

An example of acceptable collaboration includes the following:

- When authorised by the professor, discussing the issues and underlying factors of a case with fellow students, and then each of the students writing up their submissions individually, from start to finish.

UNIVERSITY STATEMENTS

ACCESSIBILITY

The Advising and Access Centre serves as Dalhousie's Centre for expertise on student accessibility and accommodation. Our work is governed by Dalhousie's Student Accommodation Policy, to best support the needs of Dalhousie students. Our teams work with students who request accommodation as a result of: disability, religious obligation, an experienced barrier related to any other characteristic protected under Canadian Human Rights legislation.

STUDENT CODE OF CONDUCT

Everyone at Dalhousie is expected to treat others with dignity and respect. The Code of Student Conduct allows Dalhousie to take disciplinary action if students don't follow this community expectation. When appropriate, violations of the code can be resolved in a reasonable and informal manner—perhaps through a restorative justice process. If an informal resolution can't be reached, or would be inappropriate, procedures exist for formal dispute resolution.

DIVERSITY AND INCLUSION

Every person at Dalhousie has a right to be respected and safe. We believe inclusiveness is fundamental to education. We stand for equality.

Dalhousie is strengthened in our diversity. We are a respectful and inclusive community. We are committed to being a place where everyone feels welcome and supported, which is why our Strategic Direction prioritizes fostering a culture of diversity and inclusiveness (Strategic Priority 5.2).

RECOGNITION OF MI'KMAQ TERRITORY

Dalhousie University would like to acknowledge that the University is on Traditional Mi'kmaq Territory.

The Elders in Residence program provides students with access to First Nations elders for guidance, counsel and support. Visit the office in the McCain Building (room 3037) or contact the programs at elders@dal.ca or 902-494-6803 (leave a message).

COURSE SCHEDULE

Date of Class	Topics	Required Readings and Material
Week of Jan 6 th	Hello World: A Course Introduction	<u>Required Readings</u> : Sweigart Chapters 0 and 1. <u>Weekly Lab Exercise</u> : Hello python world!
Week of Jan 11 th	Flow Control and Functions	<u>Required Readings</u> : Sweigart Chapters 2 and 3. <u>Weekly Lab Exercise</u> : A function for validating Elections Nova Scotia's records
Week of Jan 18 th	Basic Data Structures: Lists and Dictionaries	<u>Required Readings</u> : Sweigart Chapters 4 and 5. <u>Weekly Lab Exercise</u> : Basic data cleaning and analysis of Halifax's housing data
Week of Jan 25 th	Accessing Data Sources	<u>Required Readings</u> : Sweigart Chapters 12 and 16. <u>Recommended Readings</u> : Sweigart Chapters 6, 7 and 9. <u>Weekly Lab Exercise</u> : Analyze Halifax's housing situation using CSV data and APIs
Week of Feb 1 st	Wrangling Data	<u>Required Readings</u> : 10 minutes to pandas. <i>Pandas User Guide</i> . <u>Weekly Lab Exercise</u> : Making Airbnb data big(ger) and eas(ier) with data frames
Week of Feb 8 th	Data Visualization	<u>Required Readings</u> : Visualization. <i>Pandas User Guide</i> . <u>Weekly Lab Exercise</u> : Visually analyze iPhone app downloads
Week of Feb 15 th – WINTER STUDY BREAK – NO CLASS		

Week of Feb 22 nd	Basic Stats in Python	<u>Required Reading:</u> Khan Academy (2021). The Idea of Significance Tests. <u>Weekly Lab Exercise:</u> Do people have different attitudes towards online dating during COVID-19?
Week of Mar 1 st	Data Mining Pt. 1	<u>Required Reading:</u> Chen et al. (2012). Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining. <u>Weekly Lab Exercise:</u> Discover associations between e-commerce purchases
Week of Mar 8 th	Data Mining Pt. 2	<u>Required Reading:</u> Davidson, J. (2018). No, machine learning is not glorified statistics. Grimson, E. (2017). Introduction to machine learning. Scikit Learn (2021). An introduction to machine learning with scikit-learn. <u>Weekly Lab Exercise:</u> Predict countries of origin from e-commerce purchases.
Week of Mar 15 th	Navigating SQL with Python	<u>Required Reading:</u> Khan Academy (2021). Intro to SQL: Querying and managing data. <u>Weekly Lab Exercise:</u> Create and manage your digital book collection
Week of Mar 22 nd	Text and Social Media Analytics	<u>Required Reading:</u> Gruzd et al. (2018). Examining government cross-platform engagement in social media: Instagram vs Twitter and the big lift project. <u>Weekly Lab Exercise:</u> Identify Halifax's social media influencers
Week of Mar 29 th	The R Programming Language	<u>Required Reading:</u> None <u>Lab:</u> Analyze Canada's greenhouse gas emissions using R