

# JSC270H1S: Data Science I

## Syllabus

Winter 2021

<https://jsc270.github.io/syllabus.html>

## Teaching Team

- Professor: [Anna Goldenberg](mailto:anna.goldenberg@utoronto.ca) ([anna.goldenberg@utoronto.ca](mailto:anna.goldenberg@utoronto.ca)).
- Teaching Assistants:
  - Lauren Erdman ([lee2@cs.toronto.edu](mailto:lee2@cs.toronto.edu))
  - Matthew Edwards ([mr.edwards@utoronto.ca](mailto:mr.edwards@utoronto.ca)).

## Course Communication

- Questions about this course should be directed to the [Discourse class discussion board](#), which students can access by logging in with their UTORID. For general questions relating to course administration or content, please use a public post so the entire class can benefit from the response. If your question involves your own solution to an assignment (or your group's solution), please post a private message to ensure academic integrity. The instructor or one of the TAs will try to answer promptly. You also have the option to communicate using the Quercus discussion section.
- If you don't think your question is appropriate for the class discussion board, please reach out to the teaching team using the emails listed above. Your message should be sent from your official university email account. Please also include 'JSC270' in the email subject line, and include your full name, student ID number, and UTOR ID in the body of the email.

---

## Office Hours

- Professor: Mondays 2-3pm, or by appointment

- TA office hours
  - Lauren: Thurs, 4pm-5pm EST (Zoom link [here](#))
  - Matt: Tues, 9am-10am EST (Zoom link [here](#))

Instructor Office hours will be held on the Lecture Zoom link.

---

## Class Times and Location

### *Lectures*

- Monday, 1-3pm EST
- Lectures will be held via Zoom. The link is provided on [Quercus](#)
- The first hour will be dedicated to teaching new material, while the second hour will be used to address student questions.
- All lecture slides will be posted to the course website once the lecture is finished.

### *Laboratory Sessions*

- Wednesday, 12-2pm EST
  - Lab Sessions will also be held via Zoom. The link is provided on [Quercus](#)
  - All code will be posted to the course website once the lab is finished.
- 

## Course Description

Data Science is a relatively new interdisciplinary field that encompasses statistical methods, the computational aspects of carrying out a data analysis, including acquisition, management, and analysis of data, and the communication of analysis processes and extracted knowledge. Statistical reasoning, computing with data, and visualization play important roles in this emerging discipline. The purpose of this course is to give you a broad introduction to many of the ways data scientists learn from data, including statistical reasoning, computation and

communication. We will use the Python programming language. Tutorial labs will give students hands-on experience in executing and communicating data science problems and solutions.

Through this course, you will gain experience working on data science projects that involve using data from industry, science, or the humanities to help answer salient questions; interact with data scientists, researchers, or other professionals from academia or industry; learn how to translate data science skills across domains and think critically about data and models of data; develop strong oral and written communication skills and the ability to work in multidisciplinary teams.

---

## Learning objectives

By the end of this course, you should be able to:

- Apply and evaluate statistical methods to develop solutions to questions based on real data.
- Become proficient in state-of-the-art software packages for handling data.
- Write computer programs to wrangle and analyse data.
- Understand ethical issues related to data analysis and software development.
- Identify and answer questions that involve applying statistical methods or machine learning algorithms to complex data.
- Work in a team to solve data science problems.
- Present the results and limitations of a data analysis at appropriate technical levels for the intended audience.

---

## Evaluation

Students will be evaluated according to [University Assessment and Grading Practices Policy](#). The table below shows the weight of each assessment.

Assessment	Weight
Assignment 1	10%
Assignment 2	15%
Assignment 3	20%
Assignment 4	20%
Perusal Papers (6)	12%
Reflection Quizzes (6)	12%
Presentations (2)	10%
Universal Income	1%

---

## Course Websites

- Class slides, notes, and other important information can be found on the [course website](#).
- Questions and Answers can be posted on the [course discussion board](#)

---

## Computing

The course will use [Python 3](#) for computing.

Assignments and Labs will use [Google Colab](#), an interactive coding environment that is completely self-contained.

A similar system called Jupyter is available on <https://www.teach.cs.toronto.edu>.

[Click here for direct access to Jupyter](#) on teach.cs.

---

## Perusall

Perusall is an online tool that allows students to read and annotate course content. To get access to Perusall, please use the following link:

Perusall: <https://app.perusall.com/home>

You will need to sign up for a free account. Once your account is confirmed, you can enrol in this class using the course code: GOLDENBERG-D6NLQ.

Throughout the semester, several (6) foundational papers will be posted for students to read and analyze.

---

## Reflection Quizzes

These are short answer quizzes designed to test your understanding of the material, and to make you think critically about concepts learned in lecture.

These quizzes will be held every other week (opposite the Perusal papers), and will be completed and submitted in Quercus, under the `Assignments` tab (See this [link](#)). These assignments are to be completed individually.

---

## Course Policies

### Missed Work Policy

Valid reasons for missing an assessment include: illness; injury; or other relevant personal issues. Any of the following types of documentation will be accepted to verify a student's reason for missing an assessment:

- [University of Toronto Verification of Student Illness or Injury form](#). The form must indicate that the degree of incapacitation on academic functioning is moderate, serious, or severe in order to be considered a valid medical reason for missing.
- Student Health or Disability Related Certificate.
- A College Registrar's Letter.
- Accessibility Services Letter.

If an assignment due date is missed for a valid reason then your assignment will not be subject to a late penalty.

Other reasons for missing an assignment due date, without documentation, will require prior approval by your instructor. If prior approval is not received and an assessment is not submitted on time then your assessment will be subject to a late penalty (see [Late Penalty](#)).

## Late Penalty

The late penalty for a missed due date is 20% per day (i.e., 24 hours). For example, if the work is submitted after 5 days (including weekend days and holidays) then you will receive a grade of zero for the missed work.

## Marking Concerns

Any requests to have your work remarked must contain a written justification for consideration. Remarking requests should be made within one week of receiving your marked work. Re-evaluation appeals are at the discretion of the instructors. Note that adjustments in marks will be rare and could equally result in a lowering or raising of the mark. If a re-evaluation is completed by the instructors, the student must accept the resulting mark as the new mark, whether it goes up or down or remains the same. When appealing a re-evaluation decision, the student accepts this condition.

## Getting Help

This term you will have the option to use Discourse for class discussion. If you decide not to use Discourse it will not disadvantage you in any way, and will not affect official University

outcomes (e.g., grades and learning opportunities). If you choose not to opt-into Discourse then you can ask questions or discuss course material with the instructor or TAs during office hours.

Be sure to read [Discourse's Privacy Policy](#) and [Terms of Use](#) carefully. Take time to understand and be comfortable with what they say. They provide for substantial sharing and disclosure of your personal information held by Discourse, which affects your privacy. If you decide to participate in Discourse, only provide content that you are comfortable sharing under the terms of the Privacy Policy and Terms of Use.

You can also visit your instructor or the teaching assistants during their office hours.

## Communicating with The Professor

Questions about course material or organization, such as,

- Is it appropriate to use this statistical method?
- How do I get rid of this error message?
- What is the due date?

can be posted on the discussion forums. Questions can be posted anonymously (so that the author is anonymous to other students but not to the instructors), if desired.

If your communication is private, such as, "I missed the test because I was ill", then contact your instructor.

Always use the direct messaging functionality on [Quercus](#) to ensure that your message reaches out the instructor and/or TA's. Allow up to 72 business hours for a reply. Regular emails will not be answered.

## Academic Integrity

You are responsible for knowing the content of the [University of Toronto's Code of Behaviour on Academic Matters](#).

As a general rule, we encourage you to discuss course material with each other and ask others for advice. However, it is not permitted to share complete solutions or to directly share code for anything that is to be handed in. When an assignment is required to be completed as a team, you may share solutions and code with other members of your team, but not with another team in the class. For example, "For question 2.1 what pandas function did you use?" is a fair question; "Please show me your code for question 2.1" is not.

If you have any questions about what is or is not permitted in this course, please do not hesitate to contact [the Professor via Quercus](#).

## Accessibility Needs

Students with diverse learning styles and needs are welcome in this course. If you have an acute or ongoing disability issue or accommodation need, you should register with Accessibility Services (AS) at the beginning of the academic year by visiting <http://www.studentlife.utoronto.ca/as/new-registration>. Without registration, you will not be able to verify your situation with your instructors, and instructors will not be advised about your accommodation needs. AS will assess your situation, develop an accommodation plan with you, and support you in requesting accommodation for your course work. Remember that the process of accommodation is private: AS will not share details of your needs or condition with any instructor, and your instructors will not reveal that you are registered with AS

## Religious Accommodations

As a student at the University of Toronto, you are part of a diverse community that welcomes and includes students and faculty from a wide range of cultural and religious traditions. For my part, I will make every reasonable effort to avoid scheduling tests, examinations, or other compulsory activities on religious holy days not captured by statutory holidays. Further to University Policy, if you anticipate being absent from class or missing a major course activity (such as a test or in-class assignment) due to a religious observance, please let me know as early in the course as possible, and with sufficient notice (at least two to three weeks), so that we can work together to make alternate arrangements.

## Specific Medical Circumstances

If you become ill and it affects your ability to do your academic work, consult me right away. Normally, I will ask you for medical documentation in support of your specific medical circumstances. The University's Verification of Student Illness or Injury (VOI) form is recommended because it indicates the impact and severity of the illness, while protecting your privacy about the details of the nature of the illness. You can submit a different form (like a letter from a doctor), as long as it is an original document, and it contains the same information as the VOI. For more information, please see <http://www.illnessverification.utoronto.ca> If you get a concussion, break your hand, or suffer some other acute injury, you should register with Accessibility Services as soon as possible.

## Accommodation for Personal Reasons

There may be times when you are unable to complete course work on time due to non-medical reasons. If you have concerns, speak to me or to an advisor in your College Registrar's office;



they can help you to decide if you want to request an extension or accommodation. They may be able to provide you with a College Registrar's letter of support to give to your instructors, and importantly, connect you with other resources on campus for help with your situation.

# Schedule

## Evaluation Calendar

Written assignments are to be submitted via Quercus (please see the assignments section). The deadline for submission of assignments is 12:59pm on Mondays (right before class starts).

Number	Topic	Date Posted	
Assignment 1	Data Analysis with the COVID DELVE Dataset	Jan 11	J
Assignment 2	Exploring Student Performance with Tabular Data	Jan 25	F
A2 Presentation	xx	xx	F
Assignment 3	Modelling with Simulated Data	Feb 8	M
Assignment 4	NLP with the Twitter API	Mar 8	A
A4 Presentation	xx	xx	A

## Course Schedule

Here is our tentative schedule, which may change as the course progresses.

Date	Session	Description
Mon. Jan 11	Lecture 1	Course Introduction - What is Data Science?

<b>Date</b>	<b>Session</b>	<b>Description</b>
Wed. Jan 13	Lab 1	Data Collection and Cleaning (Intro to python/pandas)
Mon. Jan 18	Lecture 2	Exploratory Data Analysis/Visualization
Wed. Jan 20	Lab 2	Exploring and Visualizing Python Data (NumPy and Matplotlib)
Mon. Jan 25	Lecture 3	Regression: Modelling/Fit/Regularization
Wed. Jan 27	Lab 3	Scikit Learn I
Mon. Feb 1	Lecture 4	Data Simulation (Guest Lecture: Dr. Fanny Chevalier)
Wed. Feb 3	Lab 4	Version Control: Git/GitHub/Bash
Mon. Feb 8	Lecture 5	Other Classifiers/ Reproducibility (Guest Lecture: Benjamin Haibe-Kains)
Wed. Feb 10	Lab 5	Scikit-Learn II/Simulation
Mon. Feb 15	READING WEEK	xx
Wed. Feb 17	READING WEEK	xx
Mon. Feb 22	Lecture 6	Making and Evaluating Predictions
Wed. Feb 24	Lab 6	Prediction Metrics
Mon. Mar 1	Lecture 7	Unsupervised Learning/ Clustering
Wed. Mar 3	Lab 7	Unsupervised Learning in Python
Mon. Mar 8	Lecture 8	NLP I(Guest Lecture: Alistair Johnson)
Wed. Mar 10	Lab 8	Twitter API/ Text Preprocessing

<b>Date</b>	<b>Session</b>	<b>Description</b>
Mon. Mar 15	Lecture 9	NLP II: Embedding
Wed. Mar 17	Lab 9	Bag-of-words Representation/ Naive Bayes Classification
Mon. Mar 22	Lecture 10	Guest Lecture: TBD
Wed. Mar 24	Lab 10	xx
Mon. Mar 29	Lecture 11	Explainability/Interpretability/Ethics
Wed. Mar 31	Lab 11	xx
Mon. Apr 5	Lecture 12	Assignment 4 Presentations
Wed. Apr 7	Lab 12	Assignment 4 Presentations

---

## Acknowledgements

The following people have contributed to the design of the course materials and website: Prof. Fanny Chevalier (2019-2020), Prof. Nathan Taback (2019), Nicole Sultanum (2019), Prof. Anna Goldenberg (2021), Lauren Erdman (2021), Matthew Edwards (2021).