INF1340H: Programming for Data Science

**Syllabus**

**Instructor:** Shion Guha

**Teaching Assistant:** Rohith Sothilingam

**Office:** N/A

**E-mail:** [shion.guha@utoronto.ca](mailto:shion.guha@utoronto.ca) (Shion); [rohith.sothilingam@mail.utoronto.ca](mailto:rohith.sothilingam@mail.utoronto.ca) (Rohith)

**Course Web Page:** <https://q.utoronto.ca/courses/241524>

**Github Repo:** <https://github.com/shionguha/inf1340-programmingfordatascience-fa21>

**Office hours:** Wednesdays 2:00pm – 3:00pm (Shion, on Zoom); by appointment (Rohith)

**Location:** Synchronous Online

Zoom Links: (a) **Lecture:** <https://utoronto.zoom.us/j/84683050692> (637697) (b) Tutorial: (c) **Office Hours (Shion):** <https://utoronto.zoom.us/j/82011506623> (826160)

**Class Times:**

**Lecture:** Wednesdays 4:00pm – 6:00pm

**Tutorials:** Thursdays 6:30-7:30 pm and 7:30– 8:30 pm (led by Rohith)

### Official Course Description:

In this course, we use a programming language to look at the design and implementation of algorithms that are relevant to Data Science tasks. The students will look into the specifics of a programming language and extend them with libraries and exercises that will add to their background in Data Science. The exercises will deal with the analysis of data sets and alternative solutions will be discussed for their processing and evaluation of the results.

### Prerequisites: none

### Course Structure

Three (3) in-class hours per week will be divided into lectures and tutorials, in which we discuss and further probe topics covered in the lectures and readings. Note that for every one (1) hour of class time, students can expect to do 3 hours of reading and preparation work on their own, outside class.

**All coursework will be available on Quercus with detailed instructions and submission deadlines (date/time). There will also be an announcement section, which students should be responsible for checking regularly.**

### Lecture and Tutorial Topics:

* **Week 1 (Sep 15):** 
  + **Lecture:** Introduction to the course
  + **Tutorial:** Setting up Jupyter notebook on UofT/Google Collab/Anaconda/ + variable types and assignments
* **Week 2 (Sep 22):**
  + **Lecture:** Data Wrangling using pandas
  + **Tutorial:** data imports and exports using pandas
* **Week 3 (Sep 29):** 
  + **Lecture:** Data cleaning using tidy data framework (principle 1)
  + **Tutorial:** lists and arrays
* **Week 4 (Oct 6):** 
  + **Lecture:** Data cleaning using tidy data framework (principle 2)
  + **Tutorial:** conditionals (if/else)
* **Week 5 (Oct 13):** 
  + **Lecture:** Data cleaning using tidy data framework (principle 3)
  + **Tutorial:** loops (for/while/break/foreach)
* **Week 6 (Oct 20):** 
  + **Lecture:** Data cleaning using tidy data framework (principle 4)
  + **Tutorial:** loops (for/while/break/foreach)
* **Week 7 (Oct 27):** 
  + **Lecture:** Data cleaning using tidy data framework (principle 5)
  + **Tutorial:** functions (class and third party)
* **Week 8 (Nov 3):** 
  + **Lecture:** Exploratory data analysis with descriptive statistics
  + **Tutorial:** functions (user defined)
* **Week 9 (Nov 10): No Class; Fall reading week**
* **Week 10 (Nov 17):** 
  + **Lecture:** Data visualization: lines and variations
  + **Tutorial:** dictionaries
* **Week 11 (Nov 24):** 
  + **Lecture:** Data visualization: histograms and variations
  + **Tutorial:** dictionaries
* **Week 12 (Dec 1):** 
  + **Lecture:** Data visualization: complex plots and graphs
  + **Tutorial:** classes and objects
* **Week 13 (Dec 8):** 
  + **Lecture:** Conclusion to the course; next steps in HCDS; revision
  + **Tutorial:** revisions and help for final project submission

### Learning Objectives:

### Upon successful completion of this course, students will be able to:

### Be introduced to an intuitive and highly used programming language, Python

### Learn practices for writing, debugging, and testing programming code;

### Learn industry-standard principles for wrangling and cleaning data;

1. Learn ways of summarizing, describing and visualizing data;

**Relationship to Master of Information (MI) Program-Level Student Learning Outcomes:**

Master of Information Program-Level Student Learning Outcomes can be found [here](https://ischool.utoronto.ca/areas-of-study/master-of-information/).

Programming is an integral part of every Data Science project. The students in this class will be exposed to the main steps and theoretical foundations of defining the appropriate steps for performing successful programming tasks (**Outcome 1**). The practical examples of the course will combine theoretical foundations with practical approaches, such that the students can perform tasks through industry-standard frameworks (**Outcome 4**). By employing data description and visualization techniques they will be able to provide robust qualitative and quantitative interpretations of datasets. At the same time they will learn and apply the principles of providing reproducible solutions (**Outcome 5**). Finally, the course will allow students to develop their own goals and continue in life-long intellectual growth beyond graduation (**Outcome 6**).

**Class Format**

The course will consist of lectures, class discussions, and tutorials. Students are expected to attend the classes and to actively participate in the discussions and tutorials. For each class, a series of topics are provided to guide students through the readings and activities, and to frame the lectures, discussions, and studios.

Teaching and learning is a shared responsibility, influenced by individual knowledge and experience, and achieved through expanding our awareness of the different issues and approaches involved in information architecture. Commitment, preparation, and active participation are important ingredients to realize this goal. Your preparation and participation are important to your learning and the learning of your colleagues.

All the course materials will be available on the University of Toronto learning management system (Quercus) together with assignments and announcements.

**Weekly Readings:**

|  |  |
| --- | --- |
| **Week** | **Assigned Reading** |
| 2 | * Critical Questions for Big Data: Provocations for a Cultural, Technological, and Scholarly Phenomenon Link: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.441.9822&rep=rep1&type=pdf> * Engaging the ethics of data science in practice Link: <https://dl.acm.org/citation.cfm?doid=3154816.3144172> |
| 3 |  The parable of Google Flu: traps in big data analysis Link: <http://science.sciencemag.org/content/343/6176/1203?casa_token=KgrXwVt-gjAAAAAA%3AN-0zkik1A19VjMjyXD6gI8wvW-an1EkeWj9kxHKrrls1Us-z_fB9UfRrAroiM3HE8LFF-DBT7BgC_w>   HCI Across Borders Link: <https://dl.acm.org/citation.cfm?id=3108901> |
| 4 | The trouble with algorithmic decisions: An analytic road map to examine efficiency and fairness in automated and opaque decision making Link: <http://journals.sagepub.com/doi/pdf/10.1177/0162243915605575>   Obama Administration White House Report. 2016. Big Data: A Report on Algorithmic Systems, Opportunity, and Civil Rights Link: <https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf> |
| 5 |  Bias in computer systems Link: <https://vsdesign.org/publications/pdf/64_friedman.pdf>   Big data and its exclusions Link: <https://heinonline.org/HOL/Page?handle=hein.journals/slro66&div=9&g_sent=1&casa_token=&collection=journals> |
| 6 |  Economic Models of (Algorithmic) Discrimination Link: <http://www.mlandthelaw.org/papers/goodman2.pdf>   Big data's disparate impact Link: <http://www.californialawreview.org/wp-content/uploads/2016/06/2Barocas-Selbst.pdf> |
| 7 |  Auditing algorithms: Research methods for detecting discrimination on internet platforms Link: <https://pdfs.semanticscholar.org/b722/7cbd34766655dea10d0437ab10df3a127396.pdf>   Algorithmic accountability: Journalistic investigation of computational power structures Link: <https://www.tandfonline.com/doi/full/10.1080/21670811.2014.976411> |
| 8 |  Is Artificial Intelligence Permanently Inscrutable? Link: <http://nautil.us/issue/40/Learning/is-artificial-intelligence-permanently-inscrutable>   How the machine ‘thinks’: Understanding opacity in machine learning algorithms Link: <http://journals.sagepub.com/doi/pdf/10.1177/2053951715622512> |
| 9 |  The mythos of model interpretability Link: <https://arxiv.org/pdf/1606.03490.pdf>   Towards a rigorous science of interpretable machine learning Link: <https://arxiv.org/pdf/1702.08608.pdf> |
| 10 |  The scored society: due process for automated predictions Link: <https://heinonline.org/HOL/Page?handle=hein.journals/washlr89&div=4&g_sent=1&casa_token=&collection=journals>   Seeing without knowing: Limitations of the transparency ideal and its application to algorithmic accountability Link: <http://journals.sagepub.com/doi/pdf/10.1177/1461444816676645> |
| 11 |  Turkers, Scholars, "Arafat" and "Peace": Cultural Communities and Algorithmic Gold Standards. Link: <https://dl.acm.org/citation.cfm?id=2675285>   Bias and reciprocity in online reviews: Evidence from field experiments on airbnb Link: <https://dl.acm.org/citation.cfm?id=2764528> |
| 12 |  The Roots of Bias on Uber Link: <https://arxiv.org/abs/1803.08579>   Evaluating Amazon's Mechanical Turk as a tool for experimental behavioral research Link: <https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0057410> |

**Deliverables and Evaluation:**

|  |  |  |
| --- | --- | --- |
| **Evaluations** | **Due Date** | **Weight** |
| Class Performance and Discussions | Feb 16 | 10% |
| Reading Responses: Most weeks students will be asked to submit a reading response by Wednesdays 11 am EST | Wednesdays 11 am EST | 25% |
| Tutorial attendance | At scheduled tutorial times | 10% |
| Tutorial quizzes | At scheduled tutorial times | 25% |
| Mid Term Project Submission (Data Wrangling + Cleaning) | November 5 | 15% |
| Final Project Submission (Above + Descriptive Statistics + Data Visualization) | December 15 | 15% |

The course requirements and weights are final and will not be modified throughout the term. Late submissions will not be considered.

**Communication Policy:**

If you have a question, there is a high chance that other students in the course have the same question or, at least, will benefit from the answer. Please post all the questions to the INF1340 Quercus Discussion Board so everyone in the course can benefit from your questions and our answers. Students are encouraged to post answers to the questions of other students where appropriate.

Emails to the instructor and TAs must have a subject that starts with "INF1340" and include some more details, e.g., "INF1340: book appointment March 4th", **and must be submitted from your *mail.utoronto.ca* student account**.

**Readings:**

It is important to complete the required readings before the lecture in order to fully benefit from the class activities. We will be discussing them in class.

**Grading:**

Please consult the Faculty of Information’s:

* Grade Interpretation Guidelines: <http://ischool.utoronto.ca/wp-content/uploads/2020/08/grade_interpretation_revised_August2020.pdf>
* The University Assessment and Grading Practices Policy: <http://www.governingcouncil.utoronto.ca/Assets/Governing+Council+Digital+Assets/Policies/PDF/grading.pdf> (note if nothing will be handed back before the final date to drop without penalty, that should be stated in the syllabus.)
* The Guidelines on the Use of INC, SDF, & WDR: <https://www.sgs.utoronto.ca/policies-guidelines/inc-sdf-wdr/>

These documents will form the basis for grading in the course.

**Late Policy:**

There is no late policy. Late submissions are not accepted and will not be considered.

**Writing Support:**

As stated in the iSchool’s Grade Interpretation Guidelines, "work that is not well written and grammatically correct will not generally be considered eligible for a grade in the A range, regardless of its quality in other respects". With this in mind, please make use of the writing support provided to graduate students by the SGS Graduate Centre for Academic Communication (<http://www.sgs.utoronto.ca/currentstudents/Pages/English-Language-and-Writing-Support.aspx>). The services are designed to target the needs of both native and non-native speakers and all programs are free. Please consult the current workshop schedule (<http://www.sgs.utoronto.ca/currentstudents/Pages/Current-Terms-Courses.aspx>) for more information.

**Academic Integrity:**

Please consult the University’s site on Academic Integrity (<http://academicintegrity.utoronto.ca>). The iSchool has a zero-tolerance policy on plagiarism as defined in section B.I.1.(d) of the University’s Code of Behaviour on Academic Matters

(<http://www.governingcouncil.utoronto.ca/Assets/Governing+Council+Digital+Assets/Policies/PDF/ppjun011995.pdf>). You should acquaint yourself with the Code. Please review the material in Cite it Right and if you require further clarification, consult the site How Not to Plagiarize (<http://advice.writing.utoronto.ca/using-sources/how-not-to-plagiarize/>).

Cite it Right covers relevant parts of the UofT [Code of Behaviour on Academic Matters (1995)](http://www.governingcouncil.utoronto.ca/Assets/Governing+Council+Digital+Assets/Policies/PDF/ppjun011995.pdf). It is expected that all iSchool students take the Cite it Right workshop and the online quiz. Completion of the online Cite it Right quiz should be made prior to the second week of classes. To review and complete the workshop, visit the orientation portion of the iSkills site: <https://inforum.library.utoronto.ca/workshops/orientation>

The essence of academic life revolves around respect not only for the ideas of others, but also their rights to those ideas and their promulgation. It is therefore essential that all of us engaged in the life of the mind take the utmost care that the ideas and expressions of ideas of other people always be appropriately handled, and, where necessary, cited. For writing assignments, when ideas or materials of others are used, they must be cited. APA format is suggested, however you may use any formal citation format you are familiar with, as long as it is used consistently in your paper, the source material can be located and the citation verified. What is most important is that the material be cited. In any situation, if you have a question, please post it to QUERCUS. Such attention to ideas and acknowledgment of their sources is central not only to academic life, but life in general.

**Accommodations:**

Students with diverse learning styles and needs are, of course, welcome in this course. If you have a disability or a health consideration that may require accommodations, please feel free to approach Student Services and/or the Accessibility Services Office (<http://www.studentlife.utoronto.ca/as>) as soon as possible. The Accessibility Services staff are available by appointment to assess needs, provide referrals and arrange appropriate accommodations. The sooner you let them know your needs, the quicker they can assist you in achieving your learning goals in this course.

**Participation and Attendance:**

Students Discussion and interaction in the classes are an important ways to learn. Sharing your experiences and ideas with your classmates is central to your learning experience in this course. As such, you should attend and participate in every class. There will also be exercises and discussions that you will participate in within your groups in your class. Some of the activities will be very helpful in completing your assignments.

**Regrading Policy:**

This is primarily a project-based course and as such, usual re-grading policies regarding assignment submission do not apply. Students and/or groups may reach out to the instructor and TA on an ad hoc basis to inquiry about why their course

Instructors and TAs should ensure all communications with the student is in writing (e.g. follow-up e-mail) and keep a copy for later reference.

**Academic Dates**: <https://ischool.utoronto.ca/current-students/academic-resources/academic-calendar/>

**Statement of Acknowledgement of Traditional Land:**

The following is the University approved land acknowledgment statement for official ceremonies (Ceremonial Committee, Governing Council):

See: <https://www.provost.utoronto.ca/wp-content/uploads/sites/155/2018/05/Final-Report-TRC.pdf>

“I (we) would like to acknowledge this land on which the University of Toronto operates. For thousands of years it has been the traditional land of the Huron-Wendat, the Seneca, and most recently, the Mississaugas of the Credit River. Today this meeting place is still the home to many Indigenous people from across Turtle Island and we are grateful to have the opportunity to work on this land.”

See also, the Faculty of Information’s Commitment to the Findings and Call for Action of the Truth and Reconciliation Commission (approved at the Feb. 4, 2016 Faculty Council): <https://ischool.utoronto.ca/wp-content/uploads/2017/11/iSchools-TRC-Commitment.pdf>

**Equity, Diversity and Inclusion:**

The University of Toronto is committed to equity, human rights and respect for diversity. All members of the learning environment in this course should strive to create an atmosphere of mutual respect where all members of our community can express themselves, engage with each other, and respect one another’s differences. U of T does not condone discrimination or harassment against any persons or communities.

**Information about Faculty of Information iSkills and co-curricular Workshops:**

The following workshop series are exclusively available to the Faculty of Information community. Faculty of Information professors, Inforum librarians, current students, alumni, and a collective of professionals and academics from each program and concentration, work together to create these unique rosters.

Together with the MMSt and MI curricula, these academic, professional, and technical iSkills workshops provide a robust information and heritage graduate educational experience.

**iSkills Workshops**: <https://inforum.library.utoronto.ca/workshops/iSkills>

In an effort to ensure your success at the Faculty of Information, key information and skills that all Faculty of Information students must possess, regardless of program or concentration, are covered in these online orientation workshops.

**Orientation Workshops**: <https://inforum.library.utoronto.ca/workshops/orientation>

# 

# Items Specific to Remote Course Delivery

**Absence Declaration Tool**

During the COVID-19 pandemic, the University is temporarily suspending the need for a doctor’s note or medical certificate for absences from academic participation; students should use the [Absence Declaration tool on ACORN](https://www.acorn.utoronto.ca/) to declare an absence if they require consideration for missed academic work; students are responsible for contacting instructors to request the academic consideration they are seeking; students should record each day of their absence as soon as it begins, up until the day before they return to classes or other academic activities.

**FIPPA Video Recording Policy:**

This course, including your participation, will be recorded on video and will be available to students in the course for viewing remotely and after each session. Course videos and materials belong to your instructor, the University, and/or other source depending on the specific facts of each situation, and are protected by copyright. In this course, you are permitted to download session videos and materials for your own academic use, but you should not copy, share, or use them for any other purpose without the explicit permission of the instructor. For questions about recording and use of videos in which you appear please contact your instructor*.*

**Minimum Technical Requirements**

The University of Toronto has identified minimum technical requirements needed for students to access remote/online learning: <https://www.viceprovoststudents.utoronto.ca/covid-19/tech-requirements-online-learning/>

For other syllabus-related items specific to online/remote delivery see also: <https://teaching.utoronto.ca/teaching-support/course-design/developing-a-syllabus/>