

# ECO225: Big Data Tools for Economists

<b>Instructor:</b>	Prof. Khazra	<b>Email:</b> <a href="mailto:nazanin.khazra@utoronto.ca">nazanin.khazra@utoronto.ca</a>
<b>Sections:</b>	L0101 and L0201	
<b>TAs:</b>	Posted on Quercus	<b>Tutorials:</b> Posted on Quercus
<b>Course email*:</b>	<a href="mailto:eco225.khazra@utoronto.ca">eco225.khazra@utoronto.ca</a>	*shared with the TAs

Important: please check your Quercus daily and make sure notifications is on. Read the syllabus thoroughly, mark your calendars with important dates, and familiarize yourself with course rules regularly.

## 1 Course Description and Learning Outcomes

This course explores unstructured data sources such as text files, webpages, social media posts, satellite imagery, and how economists harness these types of data. It offers a practical introduction to creating datasets from these types of sources (for example, via web scrapping), linking data sources, and managing and visualizing these data (for example, via geospatial visualization).

The exercises in the course will require Python programming. Previous experience in this language is helpful but not necessary. Students that have no experience with Python can use this course as a starting point.

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

- Learn to code with Python at intermediate or fluent levels
- Learn to search effectively and debug their code
- Learn coding skills most useful to economists such as GIS mapping, web scraping, and machine learning.
- Understand the process of doing applied Economic research
- Apply their coding knowledge to a real world dataset
- Formulate a research question
- Create a full academic paper from data cleaning to visualization and results using student specific real world datasets

## 1.1 How to Succeed in This Course

ECO225 is a research-based course with no exams. To succeed in this course, it is important to adopt a different approach to learning. Here are some key things to keep in mind:

1. Make sure to work on your project consistently, *ideally every day*.
2. ***Plan to spend at least half of your time researching and seeking help to understand the various topics and tools you need for your specific research question.*** This may involve googling errors, reading documentation, and seeking assistance from the course instructors. Keep in mind that your research may differ from that of other students in your class, and you may need to learn to use additional packages depending on your chosen research question and methodologies.
3. Remember that the projects and notebook exercises will be the primary sources of learning in this course, in addition to the in-class lectures. ***ECO225 is a research course that aims to teach you not just the subject matter, but also the crucial skill of how to learn new topics on your own.*** This will become more evident as you work through the first project.
4. ***It is normal to feel overwhelmed when working on a research paper, as it can be a complex and challenging process.*** Your ideas may not always pan out as planned, and it's important to be prepared for this inherent uncertainty in research. However, you don't have to go through this process alone. Our team is here to provide support and guidance, and we encourage you to take advantage of all the resources we offer. As you work on your paper, it's important to plan ahead and stay organized. Make sure to check in with your TAs during the designated collaboration hours for additional help and support. We will discuss this in more detail later.

## 1.2 Text and Software

All textbooks and learning materials are available online for free. I use a different source for each section. Here are some useful references that we will *selectively* use in our course.

- Provided lecture notes and Jupyter notebooks
- Pro Git (Scott Chacon, Ben Straub, 2nd edition) <https://git-scm.com/book/en/v2>
- Introduction to Statistical Learning with Applications in R (James, Witten, Hastie, Tibshirani, 2013) <http://faculty.marshall.usc.edu/gareth-james/ISL/>
- Videos, slides and other material posted on Quercus
- We will mainly use Python and Jupyter notebook.

### 1.3 Prerequisites

The professor **cannot** change or wave the prerequisites. Please contact the Econ department undergraduate administrative staff if you have any questions.

Prerequisite: ECO100Y1(67%)/(ECO101H1(63%); ECO102H1(63%))/ECO105Y1(80%)/ECO100Y1(63%)/ECO102H5(63%)/(MGEA02H3 (67%); MGEA06H3 (67%)); MAT133Y1(63%)/(MAT135H1(60%); MAT136H1(60%))/ MAT137Y1(55%)/MAT157Y1(55%); CSC108H1/CSC148H1 Exclusion: ESC190H

### 1.4 Online Delivery Requirements

This course will have some online components. The lectures may be a combination of online and in-person lectures. We will announce the schedule and details on Quercus. You need high-speed internet and a laptop.

- Keep a calendar with due dates.
- All times will be posted in local Toronto time, and confusion over time zones will not be considered an appropriate excuse for missing a deadline.
- Take-home assignments are due at 7:00 pm Toronto time on the due date unless otherwise stated.

## 2 Course Rules

### 2.1 Email Policy

Before you start writing an email to a member of the course staff:

- Please make sure your question is not already answered in the syllabus or announcements on Quercus
- If this is a coding question:
  - First, try to Google the error that you get (e.g., copy and paste it into Google). Since Python is an open-source program, most of your questions have already been answered on the web.
  - If you could not fix the issue, post it on our discussion platform. Your classmates can learn from your questions. We value active participation (asking and answering questions) on our discussion platform.
  - If you still need more help, attend the collaboration hours and your TAs will answer your questions.
  - At last, if you tried all of the above and still have a question, send an email to [eco225.khazra@utoronto.ca](mailto:eco225.khazra@utoronto.ca)
- Do not reply to announcement emails
- Use your UT email address in all your communications

- Email is mainly for private communications. For content-focused questions, please use the collaboration hours. An alternative way to get answers, show participation, and benefit your classmates is to use our discussion platform.
- **Important:** please write formal emails with proper salutation, body, and closing. There are useful resources in this piece: <https://sociology.utoronto.ca/how-to-write-an-email-when-you-need-help/>.
- If you do not receive a response from me by the end of the next three business days, the most likely reason is that one of (a)-(f) above are not satisfied.

## 2.2 Technical Difficulties Policy

We will not accept missed work due to technical difficulty, deadline confusion, internet, or hardware problems. You can (but try not to) miss one weekly assignment during the semester. Please find the details in section 4; Assignments and Projects. Wisely reserve these options for unforeseen technical difficulties, illness, or other incidents.

## 3 Course Structure

### 3.1 Lectures

I will post some lectures recordings and Jupyter Notebooks on Quercus each week. **Do not share any of the course material.** *Course videos and materials belong to your instructor, the University, and/or other sources 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.

We may also have synchronous lecture sessions as needed. Dates of the live sessions will be announced on Quercus. The primary purpose of having the live lectures is to complement the posted material and, most importantly, to have in-person interactions while we are online.

I will provide the lab code and explanation for each week's material. You should run these labs each week before the lecture and use them to submit your exercises and your projects.

## 3.2 Collaboration Hours and Office Hours

We do not have lecture-based tutorials in this class. However, collaboration hours are going to have the same functionality. You should use these collaboration hours to clarify any questions that you may have about the material, exercises, and your projects.

	Time <sup>a</sup>	Held by	Appointment
Non-content consultation <sup>b</sup>	Fridays.	Instructor	Yes
Content office hours	Check Quercus	Instructor	-
Collaboration hours	Check Quercus	TAs	-

<sup>a</sup>Check Quercus for updated details on office hours and collaboration hours schedule and for the join links.

<sup>b</sup>Content questions will not be answered during non-content office hours.

## 4 Assignments and Projects

Category	Item	Weight	Due Dates <sup>a</sup>	Timing
Exercises <sup>b</sup>	Lab Exercises	22%	Most Tuesdays <sup>c</sup>	Take-home
	Term project 1	12%	Fri. Feb. 3	Take-home
Projects <sup>d</sup>	Term project 2	12%	Fri. Mar. 3	Take-home
	Term project 3	12%	Fri. Mar. 24	Take-home
	Final Project	25%	TBD	Take-home
Active Participation	Participation	15%		-
CVs	Resume Submission	2%	Jan. 17 & Mar. 28	Take-home

<sup>a</sup>All times mentioned in this table are Toronto times. Due time is at 7:00 PM of the due date.

<sup>b</sup>Your **lowest exercise mark** will be dropped. If you miss more than one exercise for any reason, they will be marked as zero. Read more in the next section.

<sup>c</sup>If class is in-person, the take-home exercise *may* be replaced by the in-class exercises and they must be submitted during class time

<sup>d</sup>If you have to miss a term project for a medical or other reasons, your **missed term project's weight** will be distributed between your other two term projects. If you miss more than one term project, it will be marked as zero. Read more in the next section.

### Special Accommodation

In case you have to miss an assignment or a project due to illness, technical difficulties, etc., you can use the special accommodation described below. You are strongly advised only to use them if necessary.

1. Your **lowest exercise mark** will be dropped. If you miss more than one exercise for any reason, they will be marked as zero. Reserve this accommodation for unforeseen circumstances.
2. If you have to miss a term project for a medical reason or other reasons, your **missed project's weight** will be distributed between your other two term projects. If you miss more than one term project, it will be marked as zero and you may consider dropping the course.

Let's work through some of the implications of this policy.

- For the first term project you miss, there is no need to self-report any reason or illness. No documentation needed/accepted. The weight of the missed project will be distributed between your two other term projects.
- If you miss a subsequent project, you may want to consider dropping the course as you will receive a zero for that project. In the case of extraordinary circumstances, contact your college's Registrar's Office. The only possibility of adjusting the marking policy would be the result of our consultation with your college's Registrar.
- You will have to include all your projects and incorporate the comments in your final project. We point out **there is absolutely no benefit to missing a term project**, even if you cannot submit a perfectly polished work.
- Further, missing a term project is risky as you do not know what the future holds. Assume that for whatever reason, you are going to be forced to miss the Final project.

#### 4.1 Resume Submission

You will need to submit your resume twice during the semester - once at the beginning and once at the end. We will provide examples and guidelines on how your resume should be structured and formatted. The first submission will be evaluated based on the quality and format of the document. For the second submission, you should include any new skills you have gained in this course and other courses you have taken this semester. The guidelines for the second submission will be posted as well. Keep in mind that you will likely need to use your resume for job applications in the near future, so it is important to make sure it is polished and professional.

#### 4.2 Participation

To succeed in this course, it is important to be an active and engaged participant. Your participation grade will be based on your attendance and involvement during collaboration hours, discussions on the online platform, and any live or in-person lectures. As this is a project-based course, you and your classmates will encounter various challenges and questions. One way to boost your participation grade is by helping your peers by answering their questions and asking your own questions on the online platform.

We will keep track of your participation in the course each week through your attendance at collaboration hours, your activity on the discussion board, and any live or in-person lectures. To ensure that you receive credit for participation, it is important that you attend collaboration hours **at least once a week** and actively participate in discussions on the online platform by asking and answering questions.

To earn participation credit in this course, it is important to:

- Attend collaboration hours at least once a week. In addition to attendance, your TAs will also evaluate your performance during these sessions, including whether you asked relevant questions or helped a classmate troubleshoot an issue. Simply attending without actively participating will not count towards your participation mark.

- Engage in discussions on the online platform at least once a week. Higher quality and thought-provoking answers will be given more weight than questions.
- Participate actively during lectures. This may include participating in polls, group work, and other activities. We will track your participation throughout the semester and award credit to those who consistently contribute high-quality participation.

I reserve the right not to disclose the distribution of the subcategories of the participation mark as they may change according to the nature of the course and, most importantly, based on the online and in-person division of the course, which is not predictable at the moment. You can calculate your participation mark after the final grades are released on ACORN given your marks for your exercises, term projects, CV submissions, and the final project. In addition to the mandatory activities, you can also show your participation in other ways by engaging in opportunities that may come up during the semester.

### 4.3 Exercises

Exercises are included in your lab for the week. You should run the lab, learn the code and concepts, and then finish each week's exercises right after.

### 4.4 Term Projects and The Final Project

You will have three term-projects and a final project. In these projects, you use the provided code and data to finish the defined tasks. We will give you detailed instructions on the steps required to complete each project. We will also provide feedback on your work, which you should then incorporate and perform the changes that we request. Make sure to address the comments you receive for each project, because you need them for your final project. In the final project, we will add some new parts, and we will also go back and check if you have incorporated our comments into your term projects.

Details about the project will be provided closer to the deadline.

### 4.5 Projects Late Submission Policy

Late Project submissions will be **penalized by day**. There is a **20% penalty** for each calendar day of late submission. For instance, if the project is due at 7:00 PM of a Wednesday, all the late submissions until Thursday 7:00 PM will incur a 20% point late penalty. There is no grace period. No submissions will be accepted five calendar days after the deadline.

### 4.6 Remarking Policy

Students should make such requests no later than two weeks after it was returned. Such a request entails a remarking of the entire work and not just the requested part. Hence, if a remarking is granted, the student must accept the resulting mark as the new mark, whether it goes up or down or remains the same. Continuing with the remark or the appeal means the student accepts this condition.

## 5 Ongoing Learning Disability or Accommodation Requirement

If you have an ongoing disability issue or you need accommodation, please register with Accessibility Services (AS) ([accessibility.utoronto.ca](http://accessibility.utoronto.ca)) at the beginning of the academic year. After AS processes your request, we will coordinate to provide the required accommodations for you. If you need accessibility related extensions, you should ask your advisor to send us the request at least one week in advance of the due date. We will then coordinate to provide the required accommodations for you.

## 6 Academic Integrity

The University of Toronto is deeply committed to the free and open exchange of ideas, and to the values of independent inquiry. As such, academic integrity is also fundamental to the University's intellectual life. What does it mean to act with academic integrity? U of T supports the International Center for Academic Integrity's definition of academic integrity as acting in all academic matters with honesty, trust, fairness, respect, responsibility, and courage.

Please visit [academicintegrity.utoronto.ca](http://academicintegrity.utoronto.ca) for smart strategies and information on academic integrity processes and procedures at the University of Toronto. The website includes a link to decisions of the University Tribunal in student cases involving academic integrity. You can review the Code of Behaviour on Academic Matters in its entirety [here](#).

Common forms of academic misconduct with code references include:

- Possession or use of unauthorized aids (B.I.1.b). Impersonation (B.I.1.c). Plagiarism (B.I.1.d) (plagiarism is a serious instance of academic misconduct, and university policy explicitly stipulates that ignorance of what constitutes plagiarism is not an acceptable defense.). Submission of work for which credit has previously been obtained (B.I.1.e). Submission of work containing purported statement(s) of fact or reference(s) to concocted sources (B.I.1.f). Assisting another student in committing an offence (B.II.1.a).

## 7 Online Etiquette

Here are a few guidelines to keep in mind as you participate in the course:

- Please do not use your personal email for any course-related communication or registration.
- Remember that there are real people with feelings on the other side of any communication or discussions. Be respectful and treat others with kindness and consideration.
- Respect the opinions of your classmates. If you disagree with or respond to their arguments, do so in a respectful manner and be sure to acknowledge any valid points they make.

## 8 Schedule and Weekly Learning Goals

The schedule is tentative and subject to change. We will try to cover as much of the material as time allows. This schedule should be viewed as a road map to the fundamental concepts that students should learn and study before each assignment.

**Week 1:** Introduction to Python, Jupyter notebook

**Week 2:** Linking Different Data Sources

**Week 3:** Version Control with Github

**Week 4:** Data Visualization

**Week 5:** Mapping with Python and GIS Mapping

**Week 6:** Satellite Data and Geospatial Visualization

**Week 7:** Satellite Data and Economic Research

**Week 8:** HTML-based Web Scraping

**Week 9:** API-based Web Scraping

**Week 10:** Working with Text

**Week 11:** Linear Regression

**Week 12:** Introduction to Machine Learning

## 9 The Final Word

As Albert Einstein once said, “A university is a place where the universality of the human experience manifests itself.”

It is crucial that we create an inclusive and welcoming environment in this class, where everyone has the opportunity to learn and grow. This is a collective effort, and I need your help to make this a reality. Let’s make an effort to connect with our classmates and embrace diverse ideas, perspectives, and identities. We are stronger and more innovative when we welcome and celebrate diversity.

I am excited to get to know all of you and work together to create a respectful and inclusive learning community.