ECO225: Big Data Tools for Economists

Instructor:	Prof. Khazra	**Email:**	nazanin.khazra@utoronto.ca
Sections:	L0101 and L0201		
TAs:	Posted on Quercus	**Tutorials:** Posted on Quercus	
Course email*:	eco225.khazra@utoronto.ca	*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. ***You will use Google and GPT a lot.*** Think of this course as a piano class; most of your learning happens when you practice at home, yet you need the instructor to guide you, teach the important core content, and refine your skills.

4. 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.

5. ***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

- Online courses and problems from https://www.datacamp.com

- 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. Please check the prerequisites on the website.

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/ChatGPT the error that you get (e.g., copy and paste it into Google/GPT). 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, use your U of T email to send an email to eco225.khazra@utoronto.ca

- 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 respectful and 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/.

- In order to maintain a positive and respectful communication environment, please be advised that *emails containing a disrespectful tone will not receive a response.* Additionally, such communication may be subject to reporting as it goes against the principles of constructive and professional discourse within our community. We appreciate your understanding and cooperation in fostering a collaborative and respectful atmosphere.

- 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 the above is 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

Please take note: It is imperative to understand that the lecture notes, project guidelines, and Jupyter notebooks *do not cover* all the topics discussed during class. *If a student misses a class, it is their obligation to catch up on the missed material.* In the event that something is mentioned in class but not included in the project guidelines, such as but not limited to the requirement to report the adjusted R squared of regressions, it will still be taken into consideration and will be graded. **Due to the nature of this in-person course, it is not possible to include every lecture detail in the project guidelines and notebooks. Part of project expectations and marking will be communicated in class and not in the guidelines.**

I will post some lecture recordings or/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.

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. Please note that signing up for office hours and not being present is an *academic integrity offense* and will be reported.

	Time	Held by	Appointment
Walk and Talk consultation[a]	Tuesday	Instructor	Yes (spreadsheet on Quercus)
Content office hours	Check Quercus	Instructor	No
Collaboration hours	Check Quercus	TAs	No

[a]Content questions will not be answered during non-content walk and talk. These hours are my gift to help my students navigate the future job market and professional obstacles and not part of my required duties.

4 Assignments and Projects

Category	Item	Weight	Due Dates[a]	Timing
Exercises[b]	Lab Exercises and/or weekly reports	10%	Mondays [c]	Take-home
Projects[d]	Term project 1	18%	Fri. Feb. 7	Take-home
	Term project 2	18%	Fri. Mar. 14	Take-home
	Final Project	14%	Apr. 4th	Take-home
	Final Paper	14%	Apr. 4th	Take-home
Presentations	First Presentation + slides	3%	Fri. Mar. 7	Take-home
	Final Presentation + slides	3%	Apr. 4th (tentative)	Take-home
Active Participation	Participation	18%		-
CVs	Resume Submission	2%	Jan. 17 & Mar. 28	Take-home

[a]All times mentioned in this table are Toronto times. Due time is at 7:00 PM of the due date.
[b]Your **lowest exercise/report 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.
[c]The take-home exercise *may* be replaced by the in-class exercises and in that case they must be submitted during class time.
[d]If you have to miss a term project for a medical or other reasons, your registrar should contact us for instructions on how to best accommodate you. If you miss more than one term project, it will be marked as zero and you may consider to drop the course. Read more in the next section.

Special Accommodation

Not that the following only applies to term projects not the final project. 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. The **lowest exercise/report 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. The **lowest participation week** will be dropped. Reserve this for unforeseen circumstances.

3. If you miss the term project's deadline due to medical or other reasons acceptable by registrar:
   1. Contact the Registrar: Email your registrar and ask them to reach out to us with instructions on accommodations (not guaranteed and varies by case).
   2. Submit ASAP: Submit your project as soon as possible, even if it's late, don't wait for the registrar's response. Waiting for the registrar's response may delay feedback on your project, which can affect your future grades (you'll need the feedback for later projects). If we grant a short extension and you miss it, penalties may apply. To avoid this, submit your work right away. Let us know if you have questions regarding this policy well-in advance of the deadline.

4. If you miss more than one term project, it will be marked as zero, and you may need to consider dropping the course. This is because there are only two term projects, and the final project builds directly upon the work completed in them.

4.1 Projects and Presentations 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 Friday, all the late submissions until Saturday 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.

4.2 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.

4.3 Participation

To do well in this course, it is essential to stay active and engaged. Your participation grade will depend on your attendance and involvement during collaboration hours, dis-

cussions on the online platform, and any live or in-person lectures. Participation carries almost as much weight as the projects because staying engaged is crucial for your success in this research-focused course, especially since there are no exams to keep you on track. Research is not something you can complete in one night-it requires consistent effort. The participation grade encourages you to stay involved with your projects daily, ensuring you learn steadily. It also helps build a supportive community, encouraging inclusion and stronger connections among students.

This project-based course will naturally present you and your classmates with various challenges and questions. A great way to improve your participation grade is by engaging with your peers-answering their questions and asking your own on the online platform.

Your participation will be tracked weekly through your attendance at collaboration hours, your contributions to the discussion board, and your involvement in live or in-person lectures. To ensure you receive full credit for participation, you must attend collaboration hours at least once per week and actively take part in online discussions by asking and answering questions.

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

- (35%) Attend collaboration hours (CH) **at least once** a week. In addition to attendance, your TAs will evaluate your performance during these sessions, including whether you asked relevant questions or helped a classmate troubleshoot an issue. Simply attending without actively participating (e.g. joining the Zoom session and leaving) will not count towards your participation mark and TAs will remove your attendance for that session. Moreover, signing up for an in-person CH and not attending may be reported as an academic integrity offense. Notes:

  - Attendance does not accumulate across weeks. For example, if you attend five CH in one week but do not attend any in subsequent weeks, you will only receive participation credit for one week, not five. Consistent weekly participation is required to earn full marks.

  - To receive full marks for the collaboration hours (CH) section, you need to attend more than one session per week. Attending one CH session earns a passing mark, while attending two CH sessions in a week earns the full mark for this subcomponent. Consistent engagement is key to maximizing your participation score.

  - CHs will start from week 2.

- (30%) Engage in discussions on the online platform at least **twice** a week. Higher quality and thought-provoking answers will be given more weight than questions. We want two posts from each student weekly (starting week 1). Answers get a higher weight than questions. You should write good answers to help your peers. Getting the helpful badge from your TA will increase your grade for this part.

- (35%) Participate actively during lectures. This may include participating in polls, group work, competitions, and other activities. We will track your participation throughout the semester and award credit to those who consistently contribute high-quality participation. **As mentioned in course structure section 3.1 above, missing class will have a negative effect on your project marks since you will miss important information about what is expected in each project.**

- Other participation opportunities will be presented during the semester. For instance, we may have a peer-to-peer feedback for presentations which will be counted in your participation mark. Or Tech-Assistant peer help roles during classes.

I reserve the right not to disclose the participation mark according to the handbook. You can calculate your participation mark after the final grades are released on ACORN given the released marks for your exercises, term projects, CV submissions, presentations, and the final project.

4.4   Exercises and Weekly Reports

You will need to submit a detailed weekly progress report based on the guidelines provided. On some weeks, this will involve completing lab exercises, which are integral to the lab content for that week. During these weeks, you should run the lab, understand the associated code and concepts, and complete the exercises. On other weeks, you will be required to submit a progress report detailing your work on ongoing projects, adhering to the specified guidelines. This structure ensures consistent engagement with course material and your projects. You can't write a research paper the week before the deadline, it is a process.

4.5   Term Projects and The Final Project

You will have two term-projects and a final project. In these projects, you use the provided sample 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 (graded). Make sure to address the comments you receive for each project, because you need them for your final project as well. 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 projects will be provided closer to each deadline.

4.6   Presentations

Students will submit two recorded presentations and their slides in the defined format. We may have peer reviews for these presentations. If we do, the peer comments will be graded as part of the participation mark in the case we have it. We will provide more details about the format of the slides, presentations, and comments during the semester.

4.7   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   Weekly Structure and Activities

Day	Main Activity	Details
Mon.	Exercise/Report (due 7PM)	Submit your exercise or report, post on Piazza, and review lab material before class.
Tues.	Lecture	2-hour lecture covering core concepts.
	Lab	10-15 minutes of a focused coding activity (coding marathon) followed by 45 minutes of lab work on your weekly report.
Wed.	CH + Project	Attend Collaboration Hours (CH) for project guidance. Work independently on your project.
Thur.	CH + Project	Attend CH, post questions or answers on Piazza, and continue project work.
Fri.	CH + Project	Attend CH for feedback or additional support. Finalize work for the week or submit your projects during proejct weeks.

- **CH (Collaboration Hours):** Dedicated times to get one-on-one or group support from the teaching team. Plan to spend 1-2 hours each week in CH. You will need more than 1-2 CHs as we move forward with the projects. So, feel free to attend as many as you need, they are free!

- **Coding Marathon:** A short, focused coding activity designed to reinforce key skills and prepare you for the weekly report.

- **Flexibility:** While this schedule is recommended, you may adapt it to fit your needs. Consistent engagement throughout the week is key.

Note: This schedule may seem overwhelming at first, but it's simply a to-do list to keep you on track and help you build good habits. Research is a steady process where consistency matters more than brilliance or big ideas. Even the best ideas can falter without proper execution, so focus on steady progress each week.

6 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) 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 <u>one week in advance of the due date</u>. We will then coordinate to provide the required accommodations for you.

7 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 for smart strategies and information on academic integrity processes and procedures at the University of Toronto. You can review the Code of Behaviour on Academic Matters in its entirety here.

Common forms of academic misconduct with code references include:

- Copy pasting text or code from Chat GPT and other AI. All words must be yours.

- 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).

Normally, students will be required to submit their course essays to the University's plagiarism detection tool for a review of textual similarity and detection of possible plagiarism. In doing so, students will allow their essays to be included as source documents in the tool's reference database, where they will be used solely for the purpose of detecting plagiarism. The terms that apply to the University's use of this tool are described on the Centre for Teaching Support & Innovation web site (https://uoft.me/pdt-faq).

7.1 Usage of GPT and AI

We will use GPT and AI in different forms in this course and you are encouraged to take advantage of them. However, copying text directly from GPT into your projects is not accepted as your work and will receive a penalty of up to 100% of your grade and may be reported. The most important part of your projects is the economic intuition that you add to your results and your analysis, and this part is one of the components that must not be generated by AI. I will explain how students can use GPT during the semester according to the status of AI at that time.

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 of the fundamental concepts that students should learn and study prior to each assignment. Please note that not all topics are used in all projects. For instance, if you are not working with very large datasets (above 100 million), you may not use SQL but it is an important tool to have in your toolbox (hence the course name) for your future job.

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: SQL (tentative/guest lecture)

Week 8: Linear Regression

Week 9: HTML-based Web Scraping

Week 10: API-based Web Scraping

Week 11: Working with Text

Week 12: Introduction to Machine Learning

9 The Final Word

Research is hard. It challenges you, pushes you, and sometimes even frustrates you. But it also teaches you something invaluable: the ability to think critically, to innovate, and to grow from feedback. Learning to take constructive feedback, even when it's tough, and using it to improve is one of the most important skills for success in any field. While it's not always easy to hear that your work needs improvement, embracing feedback without losing confidence is what sets apart those who grow and thrive.

Our teaching team is deeply committed to your success. We will carefully review your projects multiple times, providing detailed, personalized feedback. We'll meet with you weekly to guide your research and answer your questions promptly, typically within a day or two, on the discussion platform or during collaboration hours.

This course is different from most undergraduate classes. Much of your learning will happen outside the classroom, googling, debugging errors, exploring resources beyond the curriculum, and thinking deeply about your work. My role in this course is more like that of an advisor. I'll introduce you to concepts, tools, and methods, and guide you in applying them. However, you'll decide which tools are most relevant to your project and dig deeper on your own.

Here, you're not just answering exam questions, you're creating your own. You'll teach yourself what you need to know to answer those questions. While I'll demo tools and point you toward resources, you ultimately take the lead. This approach may feel unfamiliar, but it's designed to build skills that go far beyond the classroom: independence, critical thinking, adaptability, and the ability to grow through feedback.

By the end of this course, you'll have a project to be proud of, a research paper that is more than just an assignment. It will serve as a tool you can leverage for jobs, graduate school, or other opportunities. This journey is about more than learning to code or write; it's about equipping yourself with the skills to succeed in any path you choose. While the process may be challenging, the results will be deeply rewarding. You'll leave this course with confidence, resilience, and a foundation for future success.

The reason I invest so much effort into this course is simple: I believe in its impact. Over the years, I've seen students leverage the skills they've developed here to excel in top companies and universities. This course isn't just about tools or assignments; it's about equipping you to face real-world challenges with confidence. Watching you grow, succeed, and achieve your goals makes every challenge worthwhile.

I'm excited to guide you on this journey and look forward to seeing where it takes you. The work may be tough, but the rewards will be well worth it; a foundation for success in whatever path you choose!