MGT-499

Data science & causal inference for sustainability

Chiarotti Edoardo, Gallea Quentin

Cursus	Sem.	Type
Managmt, dur et tech	MA1	Obl.

Language of teaching	English
Credits	4
Withdrawal	Unauthorized
Session	Winter
Semester	Fall
Exam	During the semester
Workload	120h
Weeks	14
Hours	**4 weekly**
Courses	2 weekly
Exercises	2 weekly
Number of positions	

Il n'est pas autorisé de se retirer de cette matière après le délai d'inscription.

Remark

Courses given on UNIL Campus, open to EPFL students

Summary

This class explores key climate questions through data. Students will learn to collect, clean, and analyze data, apply causal methods using Python, and communicate insights clearly. With a focus on sustainability, the course builds skills to avoid pitfalls and draw meaningful conclusions.

Content

In today's climate change debate, pivotal questions arise: Should we stop selling fuel cars? Is a local diet healthier and more sustainable? What impact do EU environmental regulations have on emissions? Answering these questions requires more than a correlational analysis; it demands a meticulous causal examination to avoid reaching costly and misleading conclusions.
This class aims to give the students the foundations to answer such questions. They will learn how to acquire and clean data, how to describe data numerically and visually, how to use econometrics methods to causally answer a question, and how to effectively communicate their results to a broad audience. The students will be acquainted with the main pitfalls behind data analysis (Is there something else? Is it the reverse? Can we extrapolate?), and they will be equipped with the methods to overcome these challenges. Using the Python language, they will apply the concepts learned during the lectures to real-world data, with a focus on sustainability issues.

Keywords

Data Science, Sustainability, Python

Learning Prerequisites

Recommended courses

We recommend you have a knowledge in programming and Python. For an intro on Python, you can follow this notebook.

Important concepts to start the course

Before taking the class, you must be able to run a Python notebook (.ipynb) on Google Colab, and a Python script (.py) on your local computer. For the latter, you should install Anaconda, and be able to run a .py file

through both Jupyter Lab and either Spyder or PyCharm. It is also advised that you know how to use GitHub and GitHub Desktop. For local installations, you can follow this notebook. For how to use python and Jupyter, you can follow this notebook.
If you do not have a personal laptop, please reach out to us before the class starts.

Learning Outcomes

By the end of the course, the student must be able to:

- Describe the main pitfalls behind data analysis
- Investigate dataset, and the problems and bias behind the data
- Explore and clean datasets
- Visualize data
- Decide which statistical/econometrics methods to use for a given problem
- Implement these methods in Python
- Estimate model parameters from empirical observations and confidence bounds
- Synthesize their findings to a broad audience

Transversal skills

- Plan and carry out activities in a way which makes optimal use of available time and other resources.
- Demonstrate the capacity for critical thinking
- Use a work methodology appropriate to the task.
- Access and evaluate appropriate sources of information.

Teaching methods

Interactive classes, using both slides (generally 1 hour per week) and coding (generally 3 hours per week). Students are expected to bring their personal laptops. Please reach out to us before the class starts if you do not have one.

Assessment methods

The class uses a continuous evaluation system with no final exam. The evaluation consists of one group project (3-4 students per group). Students will have to apply the data science and econometrics techniques learned during the class to causally answer a question related to sustainability. The grade is made of 2 deliverables:

- Mid-term project (20%): Students will have do a short presentation to motivate their research question, present a short literature review, their samples, an initial exploratory data analysis, and discuss the potential issues they will face in their causal analysis

- Final report (80%): Students will have to write a short article to present their work, targeting a broad audience

Supervision

Office hours	No
Assistant.e.s	No
Forum	Yes

Resources

Virtual desktop infrastructure (VDI)

No

Bibliography

- [not mandatory] The Mixtape, by Scott Cunningham), Free e-book version

- [not mandatory] Python Data Science Handbook: Essential Tools for Working with Data, by Jake VanderPlas (2016), O'REILLY, EPFL library
- [not mandatory] Introduction to Computation and Programming Using Python, Revised And Expanded Edition, by John V. Guttag (2013), The MIT Press, EPFL library
- [not mandatory] A Primer on Scientific Programming with Python, by Hans Petter Langtangen (2016), Springer, EPFL library

Ressources en bibliothèque

- Find the references at the Library

Références suggérées par la bibliothèque

- Causal Inference: The Mixtape / Cunningham

Notes/Handbook

The course materials will be made available via Moodle and a Git repository

Websites

- https://github.com/edoardochiarotti/class_datascience

Moodle Link

- https://go.epfl.ch/MGT-499