

Data Science for Economists

Introduction and Motivation

Mox Ballo

University of the Philippines

January 22, 2025



SCHOOL OF ECONOMICS
UNIVERSITY OF THE PHILIPPINES
DILIMAN, QUEZON CITY

- 1 Introduction to the Course
- 2 Discussion of Syllabus
- 3 Expectation Setting
- 4 Introduction to Data Science

- 1 Introduction to the Course
- 2 Discussion of Syllabus
- 3 Expectation Setting
- 4 Introduction to Data Science

Data Science for Economists

- bridge the gap between data science and economic analysis
- provide the tools necessary to leverage data in economic research
- diverse skill set: programming, statistical analysis, and machine learning

Motivation for the Course

- Growing importance of data science in economics
- Application of these skills in research, policy-making, and industry
- Building a bridge between traditional econometrics and modern data tools.

About the Lecturer

- Economist for the Philippine Competition Commission, Business and Economics Division.
- Undergraduate degree in mathematics, graduate degree in data science, graduate units in economics.
- Specializes in geospatial data science in PCC.
- Office hours: ***By appointment.***

- 1 Introduction to the Course
- 2 Discussion of Syllabus
- 3 Expectation Setting
- 4 Introduction to Data Science

Module 1: Python Basics and Version Control (Weeks 1-2)

- *Introduction to Python* – Cover Python basics, syntax (PEP8), and its role in data science.
- *Version Control Basics* - Explain Git usage and the importance of version tracking in projects

References:

- A. E. Turrell. *Coding for Economists: A Gentle Introduction to Programming and Data Analysis in Economics* [2024].
- Guido van Rossum et al. *PEP 8: Style Guide for Python Code* [2001].

Module 2: Exploratory Data Analysis and Visualization (Weeks 3-7)

- *Data Wrangling* - Hands-on techniques for cleaning and preparing data.
- *Exploratory Data Analysis* - Identify trends, outliers, and patterns.
- *Data Visualization* - Best practices for creating impactful charts using libraries like Matplotlib and Seaborn.

References:

- Wes McKinney. *Python for Data Analysis: Data Wrangling with pandas, NumPy, and Jupyter* [3rd ed., 2022].
- Jake VanderPlas. *Python Data Science Handbook: Essential Tools for Working with Data* [2016].

Midterm Exam

Take home 36 hour exam

- includes all the covered topics in Modules 1 and 2
- notebook (*.ipynb) submission

Module 3: Data Science Applications to Econometrics

- *Introduction to Econometrics with Python* - Apply econometric models using Python, will mainly use contents from Wooldridge.
- *Introduction to Machine Learning* - Explore supervised and unsupervised learning methods.

References:

- Florian Heiss and Daniel Brunner. *Using Python for Introductory Econometrics* [1st ed., 2020].
- Gareth James et al. *An Introduction to Statistical Learning with Applications in Python* [2023].
- A. E. Turrell. *Coding for Economists: A Gentle Introduction to Programming and Data Analysis in Economics* [2024].

Module 4: Special Topics

- *Demand Estimation* - Techniques for economic modeling and market analysis.
- *Geospatial Analysis* - Tools and libraries for analyzing spatial data (e.g., GeoPandas).

References:

- Florian Heiss and Daniel Brunner. *Using Python for Introductory Econometrics* [1st ed., 2020].
- A. E. Turrell. *Coding for Economists: A Gentle Introduction to Programming and Data Analysis in Economics* [2024].

Capstone Project

- Group project (teams with 2-3 members)
- Can cover any topic.

Important Dates

- Problem Set 1 February 14
- Problem Set 2 March 14
- Problem Set 3 April 11
- Problem Set 4 May 9
- Midterm Exam March 12
- Submission of Group Members for Capstone Project March 14
- Submission of Proposed Capstone Project April 11
- Submission of Draft Capstone Project May 5
- Capstone Project Presentation May 7
- Submission of Capstone Project Write-up May 16

Course Policies

- **Learning Management System** - Will mainly use *GitHub* Classroom and students must create GitHub accounts by **January 29, 2025**
- Updates and announcements will be sent via email list and through the beadle
- **Safe Space Policy** - This will always be an inclusive and respectful environment. Please refer to *UP Gender Guidelines* for further details.

Course Policies

- **Attendance Policy:** Attendance will be checked before the 15-minute mid-class break. Full participation is expected in all activities.
- **Academic Integrity:** All students must adhere to UP's 2012 Code of Student Conduct. Cheating and plagiarism will result in disciplinary action.
- **Submission of Deliverables:** Files should be submitted in the following format: [Deliverable]_Surname.ipynb. All submissions via GitHub Classroom. *Make sure that the output are also reflected in the submitted notebook*

Course Policies

Missed Deadlines

- Notify the instructor within two days of the missed date.
- Make-up exams will only be granted for valid reasons (e.g., medical emergencies).
- Unexcused absences will result in a grade of zero for the missed activity.

- 1 Introduction to the Course
- 2 Discussion of Syllabus
- 3 Expecatation Setting**
- 4 Introduction to Data Science

Setting Expectations (Lecturer)

- **Prerequisites:**
 - Familiarity with econometrics.
 - Basic programming logic.
- **Tools:**
 - Install Anaconda and JupyterLab.
 - Ensure Python 3.10+ is installed.
 - Optionally, use Google Colab for additional RAM.
- **Engagement:**
 - Actively participate in discussions.
 - Be open to exploring new tools and methods.

What to Expect in This Course

- Hands-on projects and exercises.
- Group collaboration on a capstone project.
- Opportunities to explore cutting-edge tools.
- Learn how to:
 - Analyze datasets efficiently.
 - Create impactful visualizations.
 - Build econometric and machine learning models.

Icebreaker and Interest Survey

- **Icebreaker:**
 - Introduce yourself:
 - Name, background, and programming tools you've used (e.g., Python, R, Stata).
 - What you're looking forward to in this course.
- **Interest Survey:**
 - Do you want to learn about:
 - Web scraping? Optical Character Recognition?
 - Large Language Models (LLMs)?
 - Neural Networks (e.g., CNN, RNN)?
 - Let's shape the course together.

- 1 Introduction to the Course
- 2 Discussion of Syllabus
- 3 Expectation Setting
- 4 Introduction to Data Science

Introduction to Data Science

- **Data Science encompasses:**
 - Data collection, cleaning, analysis, and visualization.
 - Feature engineering to generate new data columns for predictions.
 - Use of machine learning and deep learning.
- **Data Science Life Cycle:**
 - Problem Identification.
 - Data Mining.
 - Cleaning.
 - Exploratory Data Analysis (EDA).
 - Feature Engineering.
 - Model Building.
 - Model Deployment.

Data Science vs Data Analytics

- **Data Science:**
 - Focuses on the entire data pipeline
 - Involves hypothesis testing and experiments.
 - Applies predictive analytics at a deeper level.
- **Data Analytics:**
 - Focuses on existing data sets.
 - Helps make decisions based on current data.
 - Limited use of machine learning.

Data Science vs Econometrics

- **Econometrics:**
 - Grounded in economic theory and models.
 - Focuses on causality and inference.
 - Seeks to understand mechanisms driving phenomena.
- **Data Science:**
 - Broader data types (structured/unstructured).
 - Uses machine learning and pattern recognition.
 - Focuses on prediction and decision-making.

Data Science vs Data Engineering

- **Data Engineering:**
 - Focuses on data platforms and storage.
 - Builds systems for data collection and management.
 - Provides infrastructure for data analysis.
- **Data Science:**
 - Uses data to generate insights and models.
 - Relies on the architecture built by data engineers.

Next Class Assignments

- **Install the following tools:**
 - Anaconda
 - JupyterLab
 - Git
- **Create a GitHub account:**
 - Sign up at <https://github.com>.
 - Familiarize yourself with the GitHub interface.
 - Submit the email address used for GitHub to the class beadle.