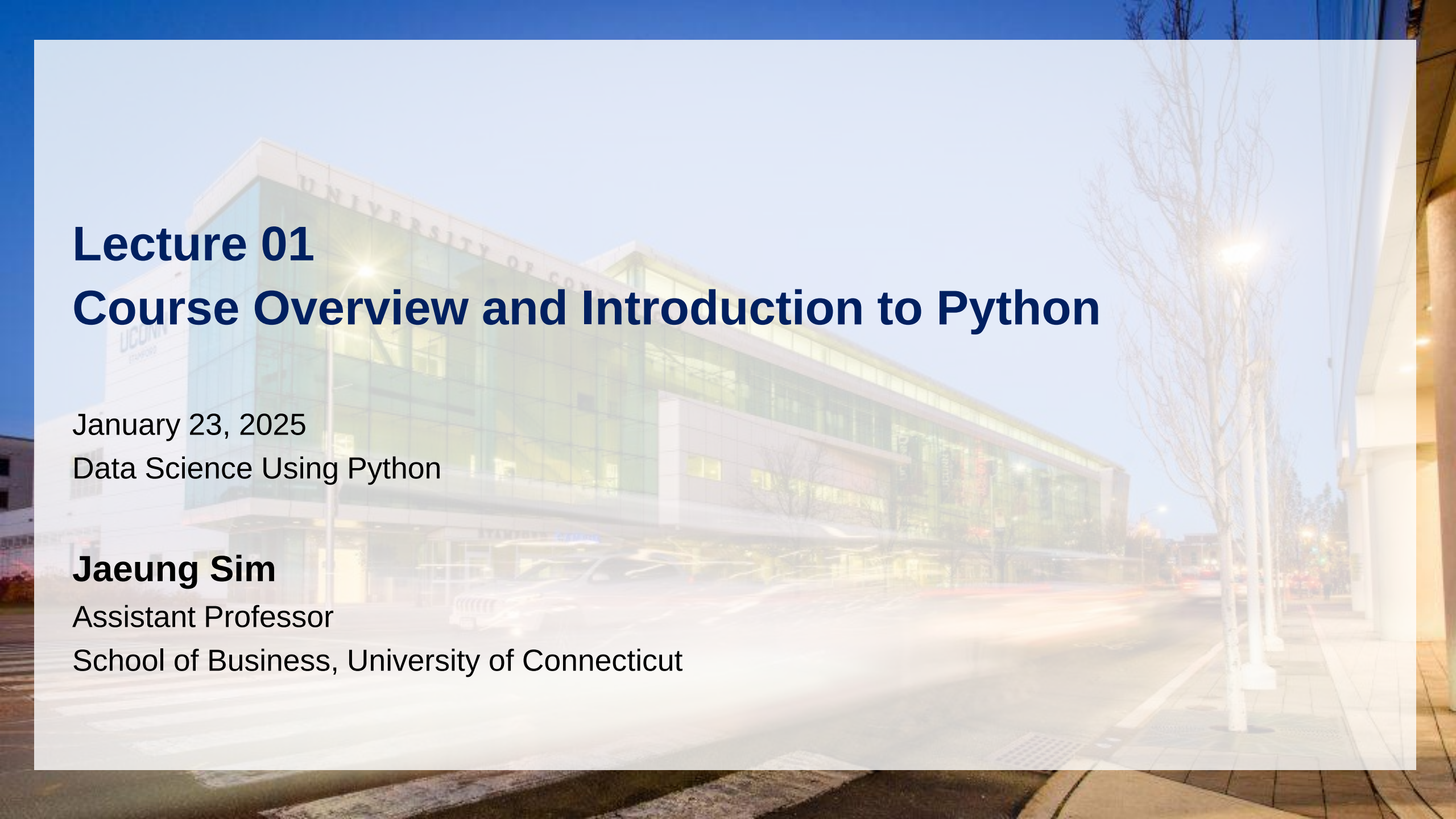


Lecture 01
Course Overview and Introduction to Python

January 23, 2025

Data Science Using Python

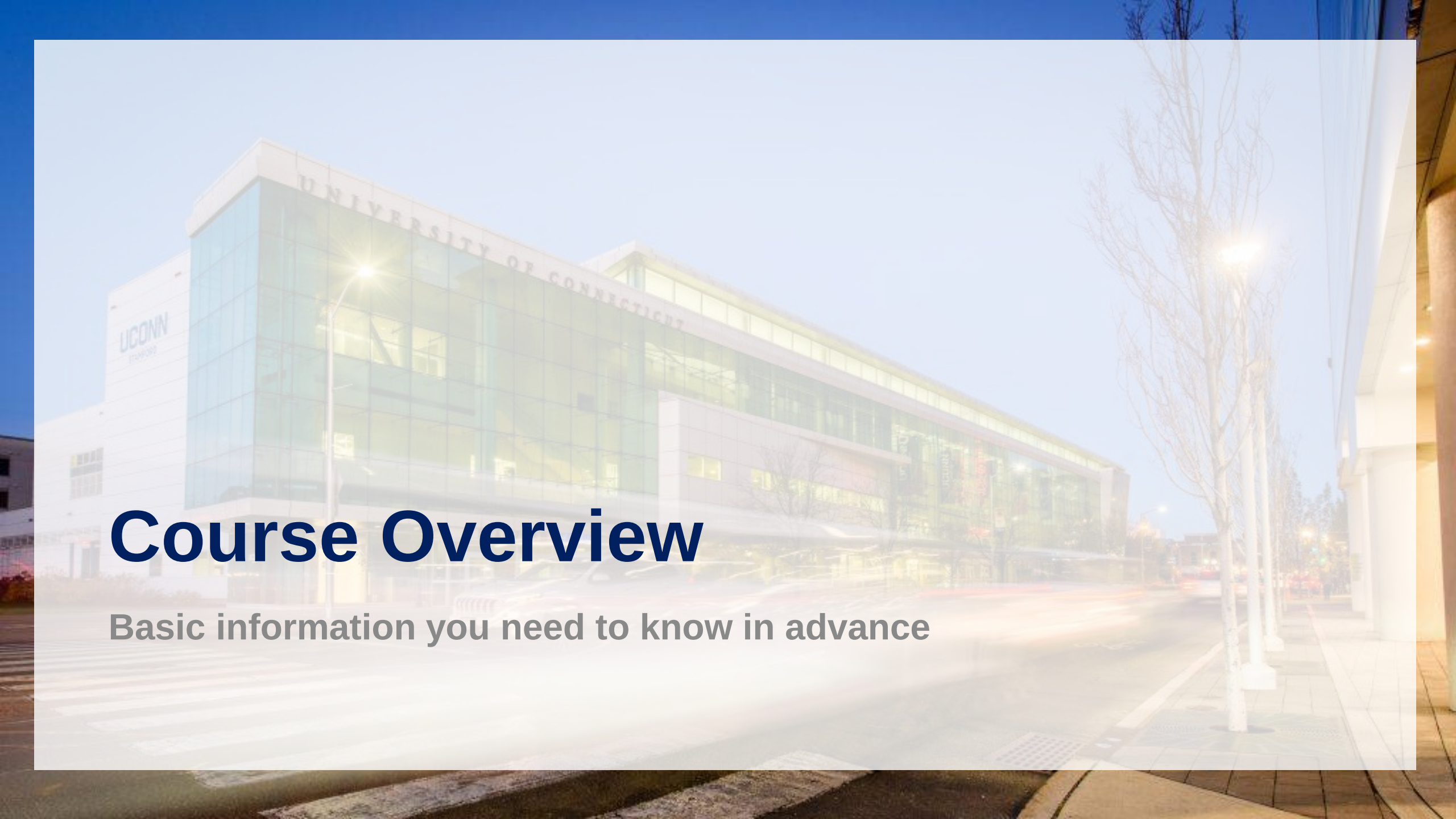
Jaeung Sim

Assistant Professor

School of Business, University of Connecticut

Agenda

- **Course Overview**
  - **Instructor**
  - Course Information
  - Course Schedule
  - Grading Schemes
  - Communication Policy
  - Academic Integrity, Responsibilities, and Rights
- **Extra Homework**
- **Warming Up!**
- **Announcements**

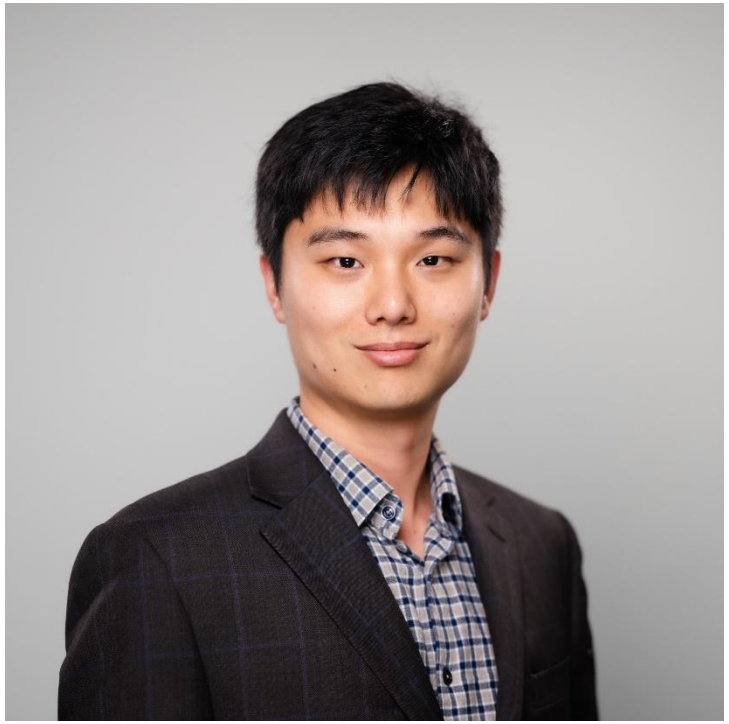


Course Overview

Basic information you need to know in advance

Instructor

- **Sim, Jaeung** [ʃim, dʒæuŋ]
  - Korean Name: **심재웅**
  - Chinese Characters: 沈載雄



Assistant Professor (Aug 2022 – present)
School of Business, University of Connecticut

Ph.D. in Management Engineering (Aug 2022)
College of Business, KAIST, Seoul, South Korea

B.S. in Industrial & Management Engineering (Aug 2015)
with *Summa Cum Laude*
POSTECH, Pohang, South Korea

Expertise (substantive)
Attention economy, platform policy, energy economics

Expertise (methodology)
Econometrics, analytical modeling, image/text mining, online/field experiments

Instructor

- **Academic Publications**
  - **Business Journals**
    - **Jaeung Sim**, Kyungmin Choi, Sang Pil Han, Daegon Cho “In-Consumption Information Cues and Online Video Consumption,” ***MIS Quarterly*** 48(2), pp. 645-678.
    - **Jaeung Sim**, Junyeong Lee, Daegon Cho (2023) “On the Effectiveness of Smart Metering Technology Adoption: Evidence from the National Rollout in the United Kingdom,” ***Journal of the Association for Information Systems*** 24(2), pp. 555-591.
    - Jae Yeon Kim, **Jaeung Sim**, Daegon Cho (2023) “Identity and Status: When Counterspeech Increases Hate Speech Reporting and Why,” ***Information Systems Frontiers*** 25, pp. 1683–1694.
    - **Jaeung Sim**, Daegon Cho, Youngdeok Hwang, Rahul Telang (2022) “Frontiers: Virus Shook the Streaming Star: Estimating the COVID-19 Impact on Music Consumption,” ***Marketing Science*** 41(1), pp. 19-32.
  - **Economics Journals**
    - Sosung Baik, Jeffrey F. Hines, **Jaeung Sim*** (2023) “Racial disparities in the energy burden beyond socio-economic inequality,” ***Energy Economics*** 127(Part A), 107098. (*Corresponding author).
    - **Jaeung Sim**, Jea Gon Park, Daegon Cho, Michael D. Smith, Jaemin Jung (2022) “Bestseller Lists and Product Discovery in the Subscription-based Market: Evidence from Music Streaming,” ***Journal of Economic Behavior & Organization*** 194, pp. 550-567.

Instructor

- **[Ongoing Research]** “Information, Emotion, or Nothing: Commentator Speech and Real-time Engagement in E-sports” (with Han Yue, Jane Gu, Hongfu Liu)

Sport & Activity

A Beginner's Guide to American Football Positions

Looking for a breakdown of all 22 players on the football field? Check out this guide to learn the basics of American football positions.

Last updated: 15 December 2023 · 12 min read

Instructor

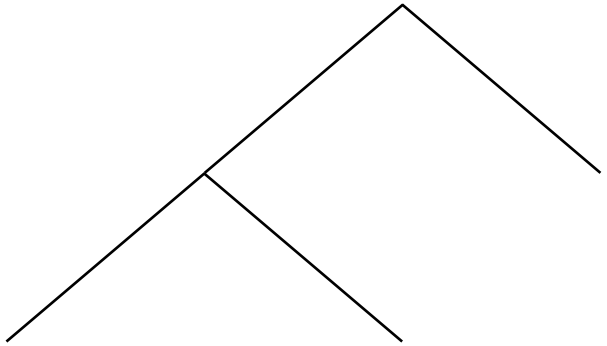
- **[Ongoing Research]** “Information, Emotion, or Nothing: Commentator Speech and Real-time Engagement in E-sports” (with Han Yue, Jane Gu, Hongfu Liu)
  - **Sports Broadcasters (Commentators)**

- **Information**
  - Description and commentary of games
- **Emotion**
  - Delivering potential emotions of teams, players, and their fans
- Bridging the gap between the event and its audience (Liu et al. 2022, Sullivan 1991).

Instructor

- **[Ongoing Research]** "Information, Emotion, or Nothing: Commentator Speech and Real-time Engagement in E-sports" (with Han Yue, Jane Gu, Hongfu Liu)
  - **Substance of Commentator Speech**

Speech Components
(by substance)



Fillers
- Having no substantive meaning
- Aiming to maintain the flow
- Heuristic processing

Information
- Describing games, players, and events
- Aiming to enhance understanding
- Systematic processing

Emotion
- Expressing subjective emotions
- Aiming to induce excitement
- Heuristic processing

Instructor

- **[Ongoing Research]** “Information, Emotion, or Nothing: Commentator Speech and Real-time Engagement in E-sports” (with Han Yue, Jane Gu, Hongfu Liu)
  - **Video analytics** (visual + audio + text) + **Econometrics** (causal inference)

Course Information

- **Class**
  - Thursday 6:15 – 9:15 pm
  - DWTN 132, Stamford (in-person)
- **Office Hours**
  - Thursday 3:30 – 4:30 pm
  - DWTN 386, Stamford (in-person)
- **Course Materials**
  - No textbooks to buy
  - Primary source: HuskyCT
  - Supplementary source: Github (https://github.com/jaeungs/uconn-opim-5512)

Course Information

- **Objectives**
  - In this course, students will learn **the basics of Python programming** and implement **advanced data science techniques** based on the programming skills.
  - This course introduces the details of Python libraries, the syntax structure and meaning, and application tips.
  - Conceptually, it focuses on predictive analysis for real-world datasets using a wide range of methods covering linear regression, deep learning, tree-based models, and reinforcement learning.

Course Information

- **Assumptions**
  - **Prerequisite: OPIM 5604 Predictive Modeling**
    - I assume that students have concrete prior knowledge of basic data analytics.
    - I do not require any technical backgrounds in statistical/programming languages prior to taking the course.

- **Technical Requirements**
  - Laptop computers for in-person classes
  - Implementing **Python** via **Google Colab** in the classroom

Course Schedule

- **First Half**

Date	Format	Topics	Notes
01/23/2025	In person	Course Overview	
01/30/2025	In person	Basics of Python Programming	HW#1 out
		Data Structure, Processing, and Visualization	
02/06/2025	In person	Data Collection Methods	
02/13/2025	In person	Regression and Predictive Analysis	
02/20/2025	In person	Deep Learning: Basics	HW#2 out
02/27/2025	In person	Deep Learning: Advanced	
03/06/2025	In person	**Mid-term Exam**	

Course Schedule

- **Second Half**

03/13/2025	In person	Deep Learning: Unstructured Data	HW#3 out
03/20/2025	N/A	*Spring Recess*	
03/27/2025	In person	Tree-based Models	
04/03/2024	In person	Reinforcement Learning: Basics	
04/10/2024	In person	Reinforcement Learning: Advanced	
04/17/2024	In person	**Term Project Presentation**	
04/24/2024	In person	Course Review & Final Exam Preview	
05/01/2024	In person	**Final Exam**	

Grading Schemes

- **Overview**

Items	Points	Weights
Class Participation	100	10%
Hands-on Assignments (x3)	300	30%
Term Project	200	20%
Mid-term Exam	200	20%
Final Exam	200	20%
Subtotal	1,000	100%

Grading Schemes

- **Class Participation** (100 points, 10%)
  - The instructor will **call your name in each class**. If you fail to respond during the class, it will affect your class participation score. Except for the first class, all students are expected to be present in the classroom. If students leave the classroom without the instructor's consent, their attendance might not be counted.
  - Virtual attendance is considered in exceptional cases, including school-approved requests for students, extreme events (e.g., hazardous weather conditions, the spread of fatal contagious diseases), and when the instructor cannot provide lectures in person due to valid reasons.
  - Students have **two tokens** that can compensate for absence. Considering the significance of missing two in-person classes, this course does not accept excuses for further absences. It is fully your responsibility to manage the risks of missing classes more than twice during the entire semester.

Grading Schemes

- **Hands-on Assignments in Python** (300 points, 30%)
  - **Hands-on Homework #1 (Data Processing and Visualization)**
    - You're going to process noisy real-world datasets and draw meaningful insights from exploring key statistics and visualizing the datasets.
  - **Hands-on Homework #2 (Predictive Model Improvement)**
    - In this assignment, you will try various model choices, including variable selections and hyperparameter tuning, and learn how they affect the predictive performance of models.
    - You might consider this assignment as **a warmup for the term project**.
  - **Hands-on Homework #3 (Image Processing)**
    - It's time to go beyond numbers and texts! You're going to extract information from images and use it to draw managerial insights.

Grading Schemes

- **Hands-on Assignments in Python** (300 points, 30%)
  - **AI use policy**
    - **University's Generatie AI Guidelines**
      - https://brand.uconn.edu/guidelines-usage/generative-ai-guidelines/
    - **Guideline in this course**
      - In your assignment, you might get help from digital resources, including online forums (e.g., Stack Overflow), generative AI tools (e.g., ChatGPT), and other online documents. However, you must do your assignment by yourself, and you are not allowed to let other people do your homework instead.

Grading Schemes

- **Term Project** (200 points, 20%)
  - This project will be a group-based prediction competition, which will be evaluated based on 1) the performance of your predictive model, 2) the validity of your modeling and programming, and 3) the presentation of your approach. Here is a brief summary of the project procedure:
    - Your team should explore model types, variable selection, and hyperparameters that are expected to show better performances in an unknown test set drawn from the same data generation function.
    - You should document a Python notebook that presents the codes and outputs in your training process and submit a spreadsheet with predicted values for the test set.
    - Your predictive performance will be evaluated based on the test set, the Python notebook will be used to validate your modeling process, and the presentation will evaluate how effectively you can communicate with others about your work.

Grading Schemes

- **Mid-term and Final Exams** (400 points in total, 40%)
  - The exam format and coverage will be announced at least two weeks before the exam date.
  - **There will be no makeup exam** unless a student is in an extremely exceptional, urgent, and verifiable circumstance.
  - Importantly, the exams aim to test both your programming skills and data science knowledge, expecting the students to digest the overall content in this course.

Communication Policy

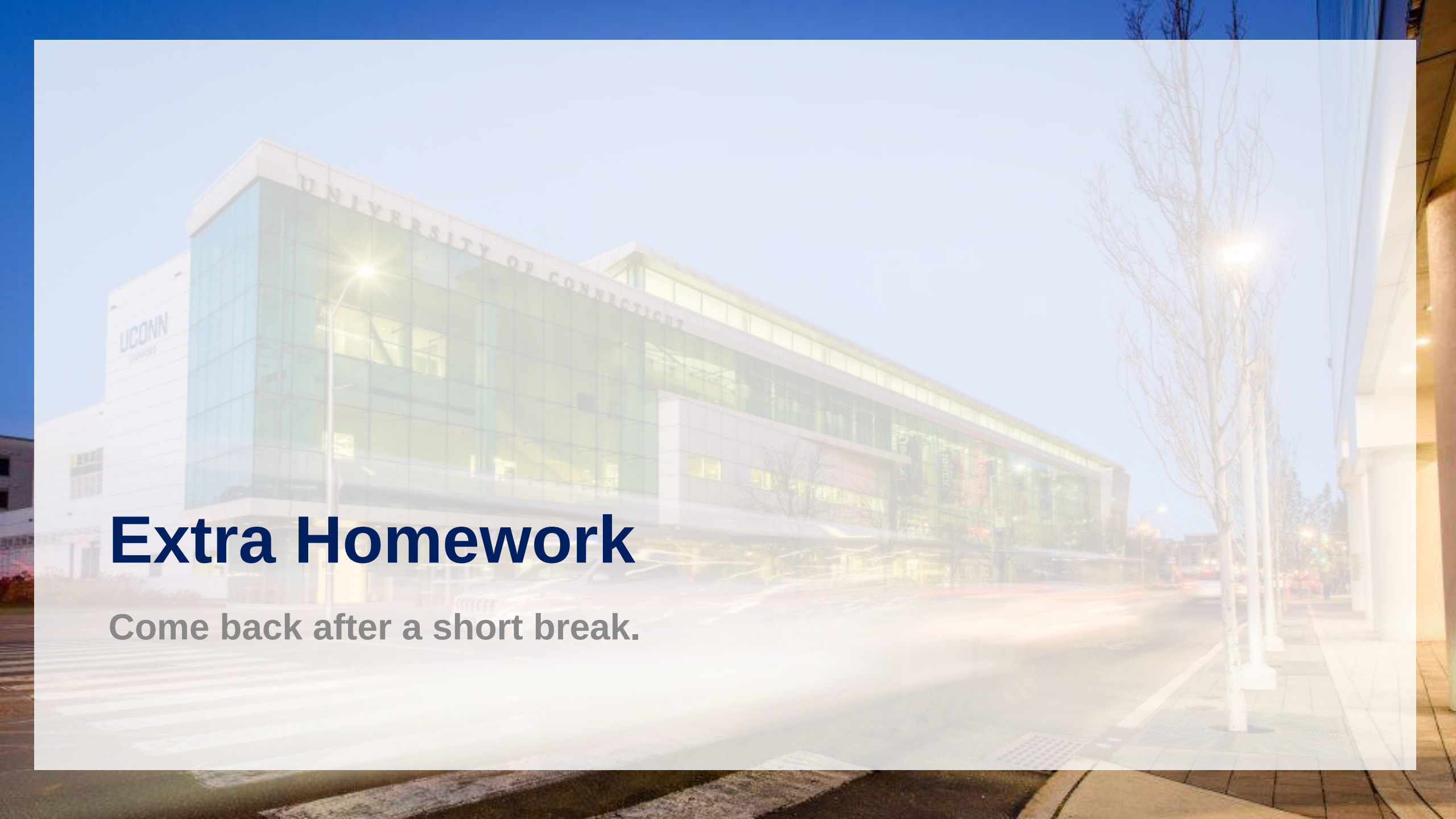
- **Language and Attitude**
  - In all types of communication, please address me politely, use appropriate language, and be concise and specific in your questions or requests.
  - Maintaining professionalism and mutual respect helps create a positive and collaborative learning environment. I encourage you to actively listen, thoughtfully engage in discussions, and approach feedback constructively.
  - Together, we can ensure effective communication that supports your academic success.
  - A supplementary slide will be shown in the classroom.

Communication Policy

- **Email**
  - I will try my best to respond to your emails typically within 24 hours, except on weekends and national holidays. In your email, please briefly describe who you are, which course you are taking, and why you are reaching out to me. If you need urgent requests or don't get a response from me over 24 hours, I strongly encourage you to send me a reminder.
- **Feedback on Assignments**
  - Before grading, I will answer clarification questions only. That is, I will not provide feedback directly related to your scores on your work, such as writing, programming codes, and slides, until I complete grading the assignment.
- **Requests for Reevaluation**
  - After you receive scores from assignments, quizzes, and the final exam, you might send me a rebuttal letter in one week. In this letter, you should clearly state for which parts and for which reasons you think your scores can change.
  - For class participation and term projects, the instructor's evaluation criteria should be strictly respected, and the scores in these parts will change only if the instructor makes an apparent mistake in the evaluation.

Academic Integrity, Responsibilities, and Rights

- Please refer to the syllabus for details.
- A supplementary slide will be shown in the classroom for academic integrity.



Extra Homework

Come back after a short break.

Introduce Yourself

- **Submit your self-introduction to HuskyCT!**
  - **Extra 10 points** will be given.
  - Please use the PowerPoint template.
  - **Deadline: Jan 29 (Wed), 2025**
  - **Students who're taking OPIM 5671 now: Please bring another fun fact!**

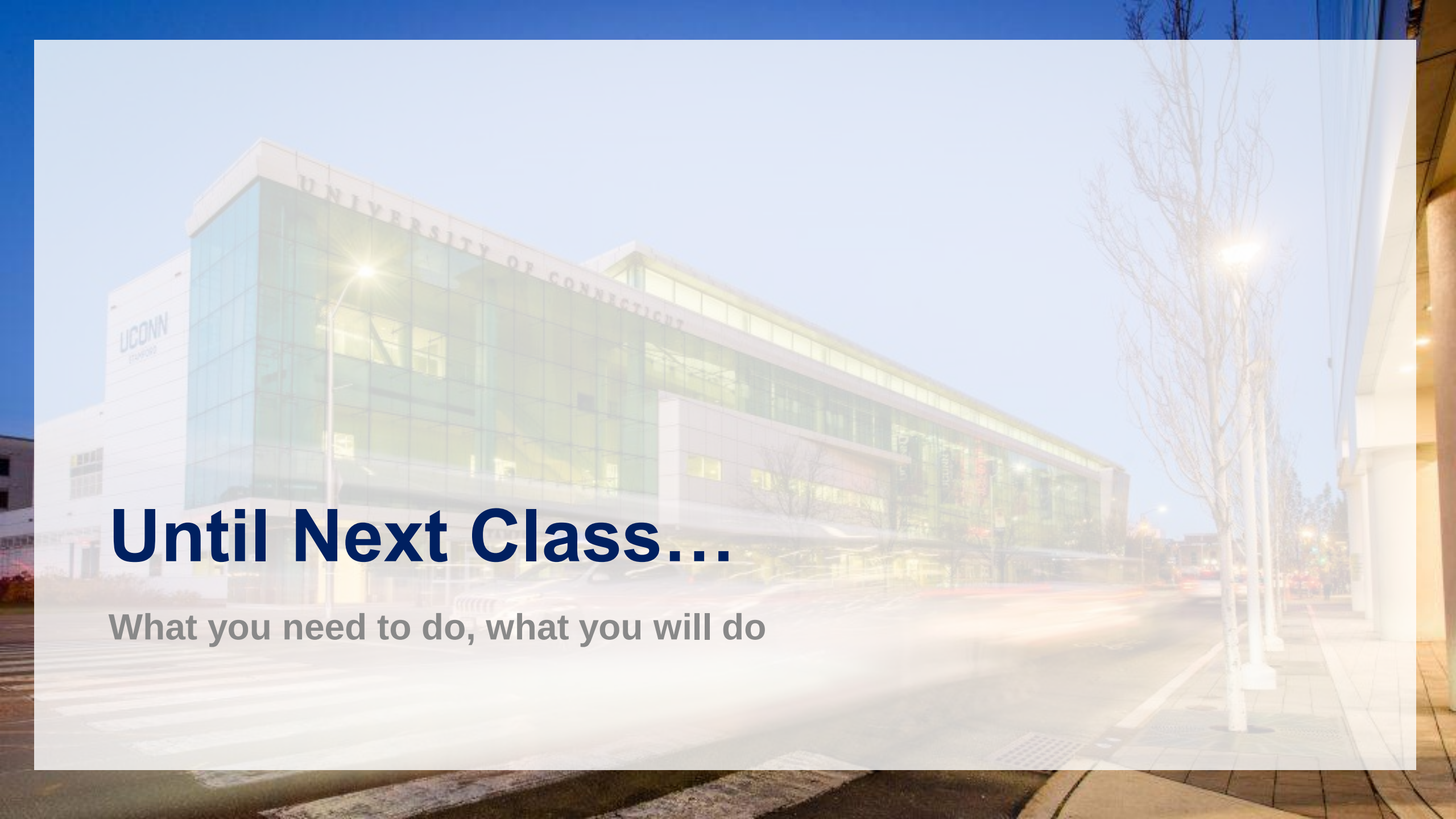
- **Self-introduction in the next class**
  - You will introduce yourself with these slides next class.
  - Minimum of two, maximum of three slides are recommended.

Warming Up!

Data Mining, Text Mining, and Time Series

Live Survey

- **Join the survey!**
  - https://pollev.com/jaeungsim088



Until Next Class…

What you need to do, what you will do

One week to our next meeting

- **Jan 30 (Thu): Next Class**
  - **Basics of Python Programming**
    - We will set up the Google Colab environment together!
    - Basic syntax, functions, and Markdown language will be covered briefly.
  - **Data Structure, Processing, and Visualization**
    - Essential libraries, such as *NumPy*, *Pandas*, and *Matplotlib* will be covered with several hands-on practices.
- **What you need to do**
  - **Submit your extra homework!**

Jaeung Sim
Assistant Professor
School of Business, University of Connecticut
jaeung.sim@uconn.edu