

# INSY336 - Data handling and *coding* for analytics 🚀

Kyunghee Lee

# About me

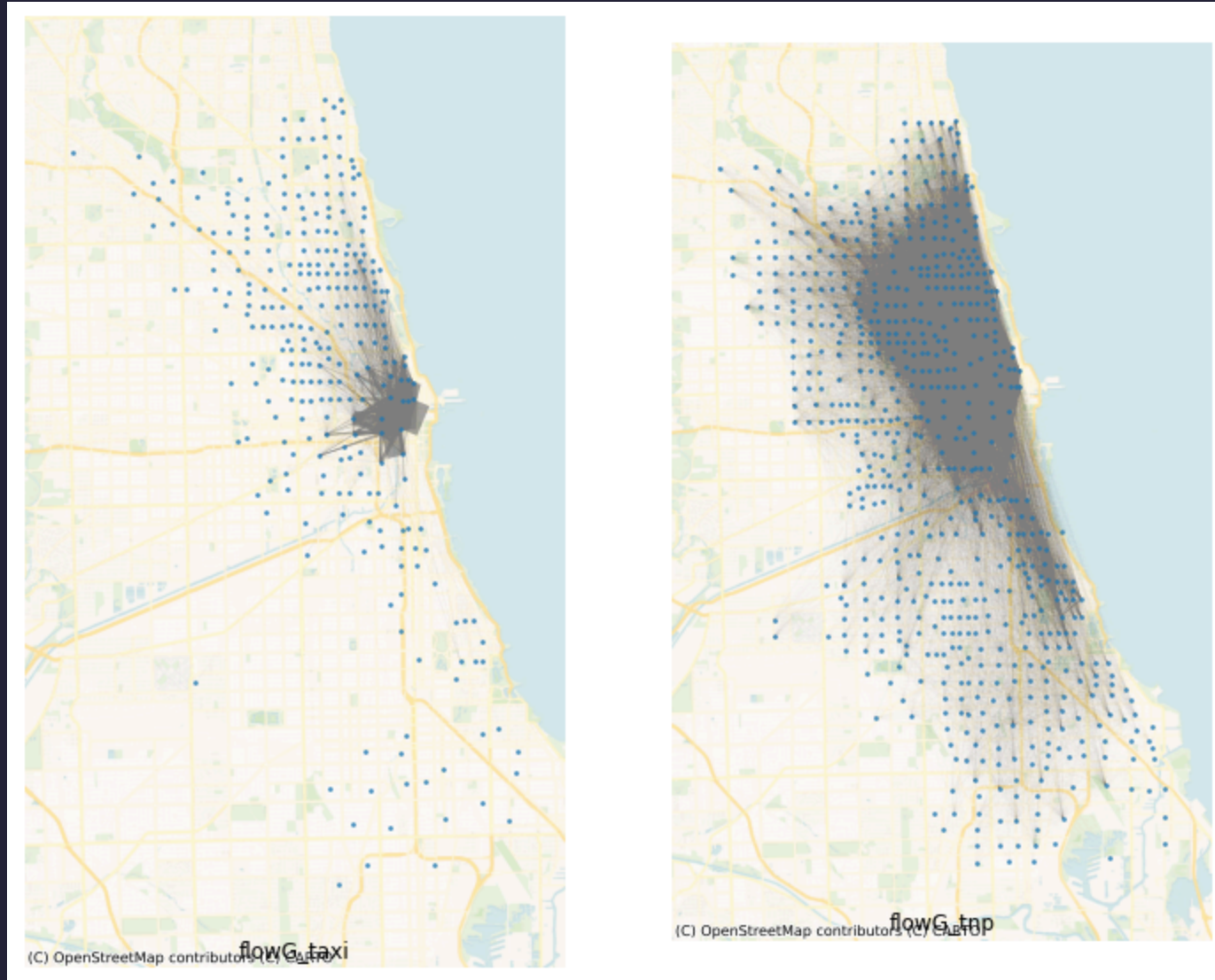
- Kyung Hee Lee
- Assistant Professor of Information Systems, McGill (2023-)
  - AP, Wayne State University, Detroit (2017-2023)
  - Postdoc, McGill (2016-2017)
- PhD in Management Engineering, KAIST, Seoul
- Research: digital platforms, decentralized applications, business value of IT



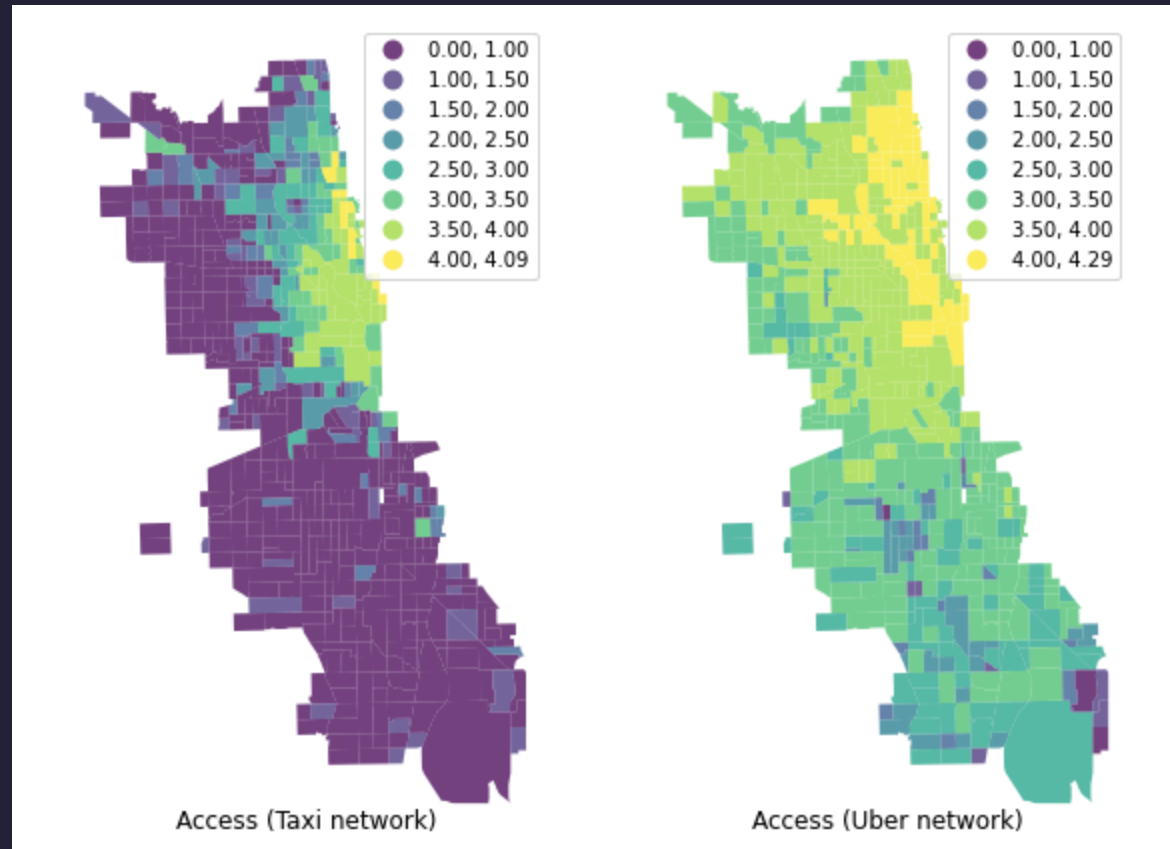
- Dugmee Hwang (Dug Mee Hwang)
- PhD student in Information Systems

**Q: Can Uber improve access to healthy foods?**

# Taxi vs. Uber trips in 2019 (from Chicago Data Portal)



# Taxi vs. Uber: Accessibility to grocery stores



# **Data Scientist: The Sexiest Job of the 21st Century**

Meet the people who can coax treasure out of messy, unstructured data. by Thomas H. Davenport and DJ Patil

From the Magazine (October 2012)

# **Is Data Scientist Still the Sexiest Job of the 21st Century?**

by Thomas H. Davenport and DJ Patil

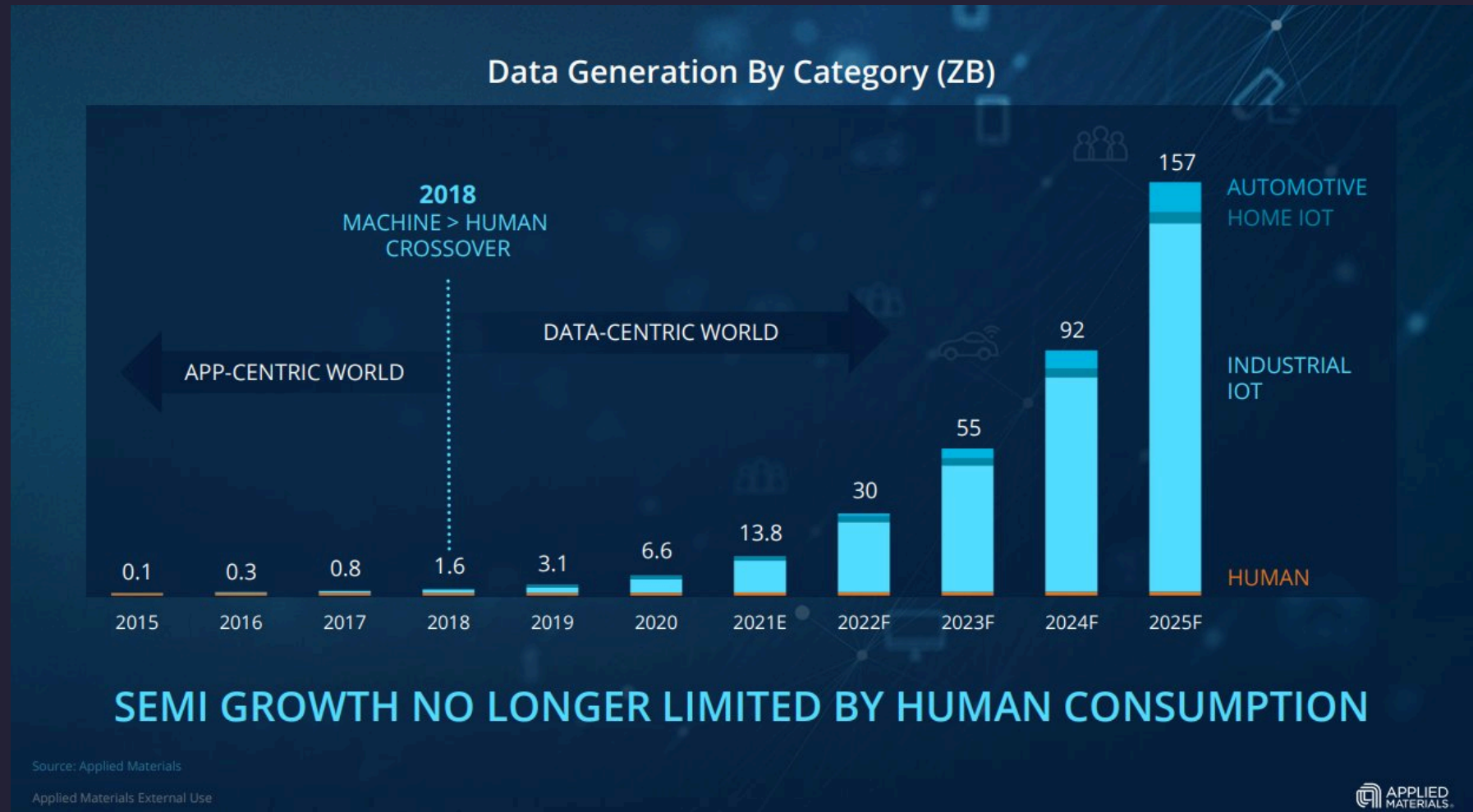
July 15, 2022

---

<https://hbr.org/2012/10/data-scientist-the-sexiest-job-of-the-21st-century>

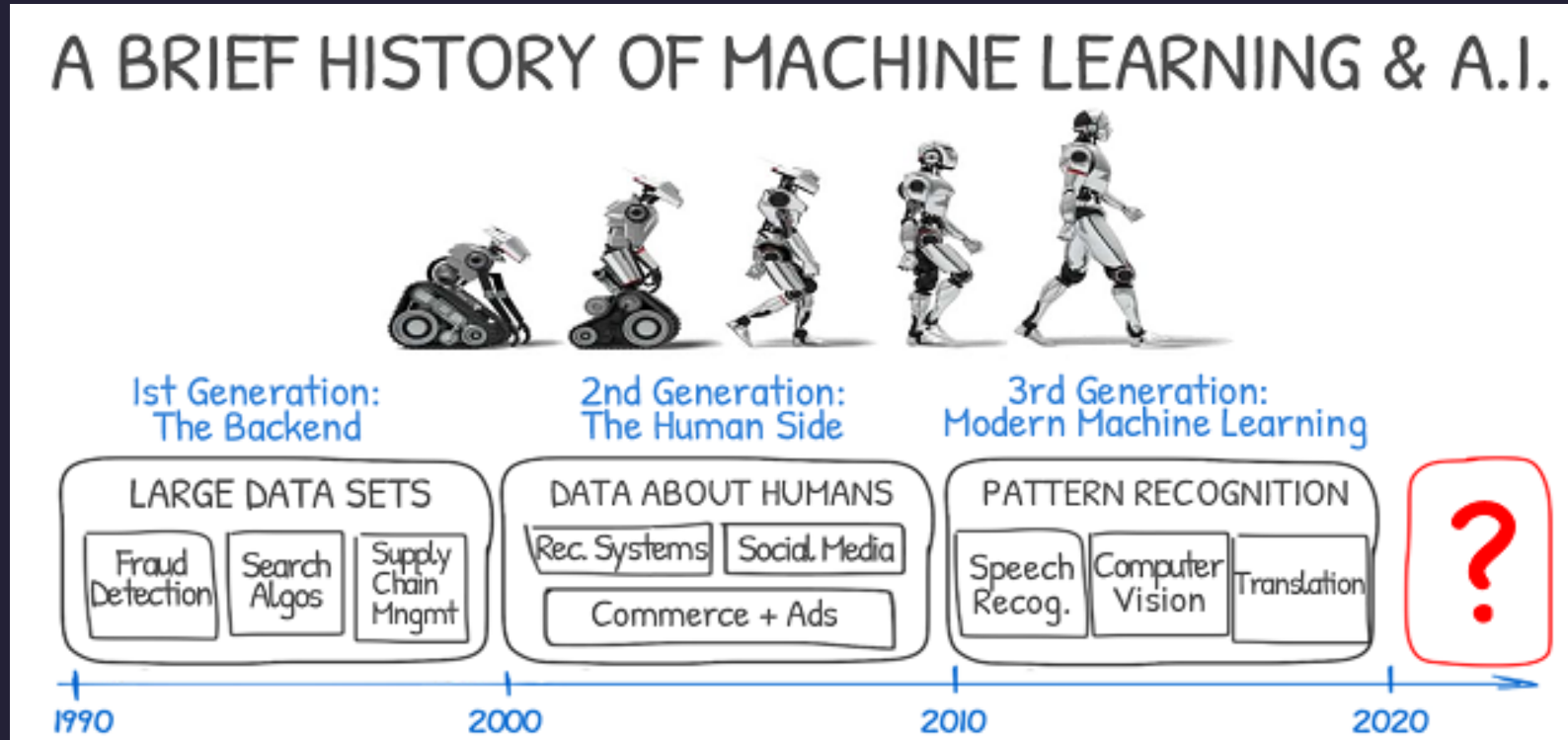
<https://hbr.org/2022/07/is-data-scientist-still-the-sexiest-job-of-the-21st-century>

# Data is everywhere





# Data is the new oil



<https://medium.com/@ODSC/the-history-and-future-of-ai-with-michael-i-jordan-85f9962ff139>

[https://ana.blogs.com/maestros/2006/11/data\\_is\\_the\\_new.html](https://ana.blogs.com/maestros/2006/11/data_is_the_new.html)

# How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did



**Kashmir Hill** Former Staff

Tech

*Welcome to The Not-So Private Parts where technology & privacy collide*

 This article is more than 8 years old.

Every time you go shopping, you share intimate details about your consumption patterns with retailers. And many of those retailers are studying those details to figure out what you like, what you need, and which coupons are most likely to make you happy.

[Target](#), for example, has figured out how to data-mine its way into your womb, to figure out whether you have a baby on the way long before you need to start buying diapers.

Charles Duhigg outlines in the [New York Times](#) how Target tries to hook parents-to-



**TARGET**

Target has got you in its aim

## Targeted Ad

- estimated "pregnancy score" based on purchase pattern
- sent coupons along the different stages of pregnancy

<https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/#6298ff1e6668>



## AI recruiter

<https://www.youtube.com/embed/8QEK7B9GUhM>

<https://www.economist.com/business/2013/04/06/robot-recruiters>

# Spatial Finance

Brokerage

## HOW SATELLITE IMAGERY IS HELPING HEDGE FUNDS OUTPERFORM

June 26, 2020

### **How hedge funds use drones, satellite images and web scraping to gain an edge**



By Matt Egan, CNN Business

Published 4:33 PM EDT, Wed July 10, 2019

# How AI Could Empower *Any* Business

[https://youtu.be/reUZRyXxUs4?si=jKVTw\\_JA4x7NxJNU](https://youtu.be/reUZRyXxUs4?si=jKVTw_JA4x7NxJNU)



# Course objective

- Understand and apply fundamental programming concepts in Python.
- Utilize Python libraries to read and manipulate data.
- Perform basic data cleaning, wrangling, and analysis tasks.
- Use SQL for managing data in relational databases.
- Implement an end-to-end Extract-Transform-Load (ETL) pipeline.



## Extract

Retrieves and verifies data  
from various sources



## Transform

Processes and organizes  
extracted data so it is usable



## Load

Moves transformed data  
to a data repository



**Module 1 - Fundamentals of programming in Python**

**Module 2 - Data Manipulation in SQL**

**Module 3 - Extract-Transform-Load (ETL) using Python and SQL**

# Logistics

Course website: <https://insy336.github.io/{section number}>

- Section 001: <https://insy336.github.io/001>
- Section 002: <https://insy336.github.io/002>

**MyCourses:** for grades

## Questions

- Course content: post on Ed Lesson
- Personal matters: email me or TA

# Tools

- Python
- SQL
- Ed Lesson (<https://edstem.org>)
- DataCamp (<https://www.datacamp.com>)

# Coursework

- In-Class Exercise (ungraded)
- Take-Home Exercise
- Assignment
- Quiz
- Exam

# Grading policy

Component	Weight (%)
Participation	5
Take-Home Exercise	10
Assignment	25
Quiz	30
Exam	30

- Late submission: 10% deduction/day; won't be accepted after 10 days
- Deferring exam: only for exceptional circumstances; documentation required
- Academic integrity
- Refer to **syllabus** for details

# Getting to know each other

- Background
- Career goal?
- What got you interested in this course?
- What do you do for fun other than coding?
- Tips for surviving Montreal winter?