Web Analytics and Intelligence

Session 1a. Introduction

Renyu (Philip) Zhang

1

Who Am I?

- I am a scholar, a teacher, and a practitioner in data science/AI and operations research.
- Research:
  - How to use data analytics to improve business decision making, especially for digitalized online platforms.
- Teaching:
  - Data science/AI for business to undergraduate, master, EMBA and PhD students.
- Data Science Practitioner:
  - Economist and Tech Lead, Kuaishou (快手; https://www.kwai.com/).
  - Evaluating and optimizing the ecosystem of Kuaishou.

- CUHK Business School, Associate Professor (*with tenure*), since 2022
- NYU Shanghai, Assistant Professor, 2016-2022; Visiting Scholar, since 2022
- Washington University in St. Louis, PhD, 2011-2016
- Peking University, BS, 2007-2011

2

2

Prof. Kevin Chen

Assistant Professor

Contact
Department of Decisions, Operations and Technology

Room 936, 9/F
Cheng Yu Tung Building
12 Chak Cheung Street
Shatin, N.T., Hong Kong

+852 3943 7902

kevinchen@cuhk.edu.hk

← Academic Staff

Chen, Kevin Hongfan（陳泓帆）

BS (Georgia Institute of Technology); MBA, PhD (Chicago Booth)

Biography

Hongfan (Kevin) Chen is an Assistant Professor of the Department of Decisions, Operations and Technology at The Chinese University of Hong Kong (CUHK) Business School. He received his PhD and MBA degrees in Operations Management from the University of Chicago Booth School of Business. Prior to Chicago Booth, he studied Industrial Engineering and Applied Mathematics at Georgie Institute of Technology. Before joining CUHK, he also had some professional experience in companies including Airbnb, Amazon, Cox Enterprises, Interface, Hewlett Packard and ABB. Professor Chen's research primarily focuses on revenue management, platform marketplaces, and optimisation under uncertainty.

Research Interests
Revenue Management
Platform Economy
Optimisation under Uncertainty

Publications & Working Papers

- J. Birge, O. Candogan, H.Chen and D. Saban (2020), "Optimal Commissions and Subscriptions in Networked Markets," *Manufacturing & Service Operations Management*.

3

3

Why are You Here?

I guess it is (not) because you find me handsome ☺

4

4

What is Web Analytics?

Web analytics is the gathering, synthesizing, and analysis of website data with the goal of improving the website user experience.

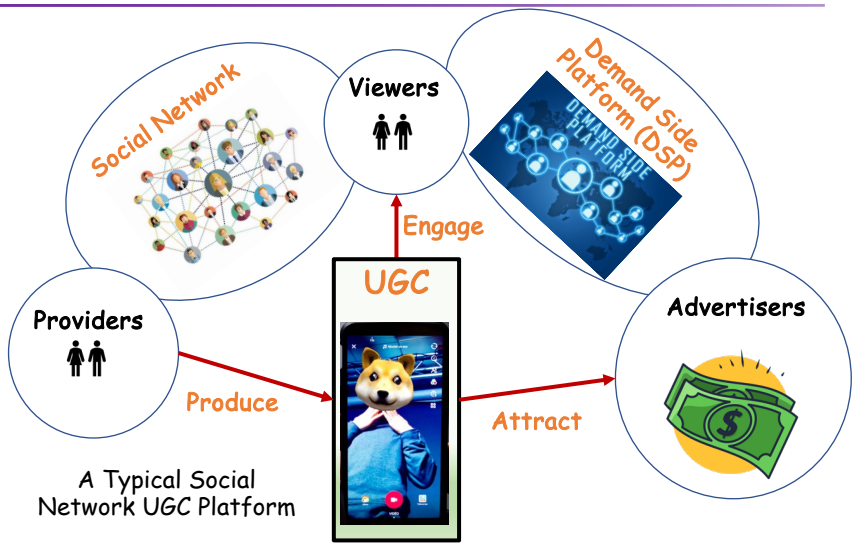
https://amplitude.com/blog/web-analytics

Conversion Funnel

5

5

What is Web Analytics?



A Typical Social Network UGC Platform

6

6

What Happened in the Past Ten Years?

- What drives the social and economical growth of the past 10 years?
  - Technology
  - Data
  - Business Model
- What should we expect next?

7

7

Most Valuable Firms in 2011

Top companies: Biggest

By Market Value | By Equity | By Employees

World GDP in 2011: $68.1 Trillion

Rank	Company	500 Rank	3/25/2011 ($ millions)
1	Exxon Mobil	2	414,638.0
2	Apple	35	323,866.1
3	Microsoft	38	215,269.0
4	Chevron	3	214,355.5
5	Berkshire Hathaway	7	210,787.5
6	General Electric	6	209,715.2
7	International Business Machines	18	197,784.3
8	Google	92	186,399.2
9	Wal-Mart Stores	1	182,764.3
10	J.P. Morgan Chase & Co.	13	182,683.8
11	AT&T	12	170,544.8
12	Procter & Gamble	26	170,511.7
13	Wells Fargo	23	168,630.3
14	Oracle	96	165,175.2
15	Pfizer	31	162,621.2
16	Johnson & Johnson	40	161,622.1
17	Coca-Cola	70	149,688.2
18	Bank of America Corp.	9	135,016.2
19	Citigroup	14	129,868.8
20	ConocoPhillips	4	116,812.3

Source: Fortune 500

8

8

Most Valuable Firms in 2023

Rank	Name		Market Cap
1	Apple AAPL	x9	$2.975 T
2	Microsoft MSFT	x13	$2.808 T
3	Saudi Aramco 2222.SR		$2.155 T
4	Alphabet (Google) GOOG	x9	$1.742 T
5	Amazon AMZN		$1.516 T
6	NVIDIA NVDA		$1.203 T
7	Meta Platforms (Facebook) META		$877.58 B
8	Berkshire Hathaway BRK-B	X3.7	$787.55 B
9	Tesla TSLA		$744.53 B
10	Eli Lilly LLY		$564.92 B

Rank	Name		Market Cap
11	Visa V		$520.97 B
12	TSMC TSM		$511.63 B
^5 13	Broadcom AVGO		$506.94 B
v1 14	UnitedHealth UNH		$502.93 B
v1 15	Novo Nordisk NVO		$460.64 B
v1 16	JPMorgan Chase JPM	x2.4	$443.27 B
v1 17	Exxon Mobil XOM		$416.67 B
v1 18	Walmart WMT		$416.30 B
19	Tencent TCEHY		$401.38 B
20	LVMH MC.PA		$390.46 B

World GDP in 2023: $105 Trillion (estimated), a 42% increase from 2011

Those tech firms' total market capitals: ~$13 Trillion, which means 1/8 of World GDP

Source: https://companiesmarketcap.com/

Some Interesting Numbers

Gross Merchant Values (GMV) of Double-Eleven at Taobao and Tmall

2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
0.052	0.936	5.2	19.1	35.0	57.1	91.2	120.7	168.2	213.5	268.4	498.2	540.3	540.3*	545.7*

Source: Alibaba Group, in Billion RMB

Monthly Active Users (MAU) of WeChat

2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023
0	158	355	500	697	889	989	1,082	1,132	1,225	1,250	1,313	1,336

Source: Tencent Group, in Million

What happened in the past decade?

Technology

- (Mobile) Internet penetration rate of China in 2023: 76.4% (60.5% for rule areas), aka 1.1 billion. These figures were 38.3% (26.4% for mobile Internet users) and 0.513 billion in 2011.
  - Source: www.ce.cn, paper.people.com.cn
- 3G-Technology (International Mobile Telecommunications-2000): 300Kbps~2Mbps
- 4G-Technology: 1Gbps+
- 5G-Technology: 20Gbps+
- Average cellphone time per day per user in 2023: 7.2h.
  - Source: QuestMobile

Technology revolution makes digitalization possible and scalable!

11

11

Data

- We cannot optimize those that cannot be measured with data.
- Digitalization keeps track of a lot of data that are otherwise not available.
- Data science and data-driven AI:
  - Image-Net, AlexNet, DNN, RNN, GAN, Alpha-Go, Alpha-Fold, GPT/LLM……
  - Facial recognitions, auto-translations, recommender systems, auto-driving……

Digitalization makes data available and accessible; data analytics, in turn, makes digitalization efficient.

12

12

Business Models

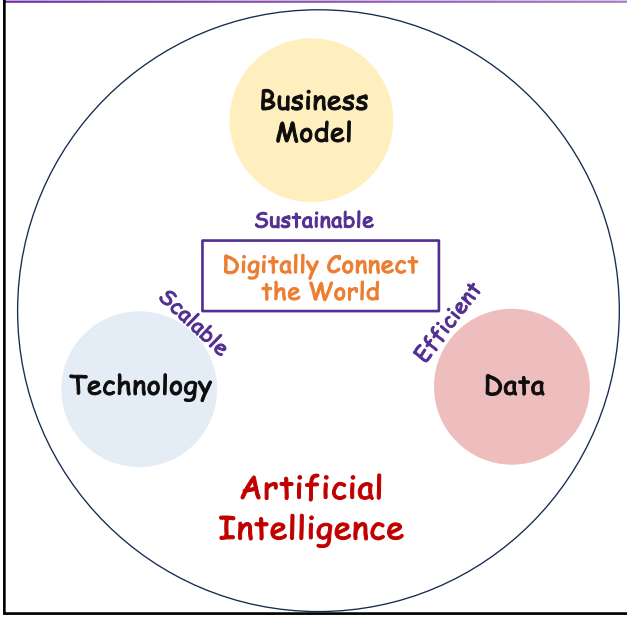
- New business models emerge as a result of digitalization.
  - Amazon, Google, Facebook, TikTok, Uber, AirBnb, Meituan, Tesla, Robinhood……
- Core business logic: Cost-effectively creating and strengthening connections with the help of digitalization.
- By 2030, the market values of major online platforms will constitute 30%+ of global annual GDP. (Source: US government)

Digitalization is the infrastructure for online platforms; platforms, in turn, make digitalization profitable and sustainable.

13

13

What's Next?



"Mostly Harmless" web analytics:

- Values of digital connections vs. Antitrust
- Efficiency from data vs. Data privacy/security
- Growth and engagement vs. Algorithmic fairness
- Automation vs. ethics

14

14

Warnings

"Big data is like teenage sex; everyone talks about it, no one really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it." ---Dan Ariely, Duke University

15

15

Topics Covered in this Course

- **Philip Zhang:**
  - Causal inference, i.e., to answer the "what if" questions, in the web analytics business contexts.

- **Kevin Chen**
  - Estimation, optimization, auctions, revenue management, etc.

16

16

Course Objective

- Our goal: Introduce the modern landscape of internet and convince you of the tremendous business (and social) value of analytics in this business sector.
- At the end of this course, hopefully, you will
  - Have a basic understanding of the internet industry;
  - Be ready to solve a real business problem in the internet sector using data and analytics tools (including the AI-powered copilots for coding);
  - Be well-prepared for an entry-level job in the internet industry.
- We will try to strike a good balance between Web and Analytics.
- Connection to real business and job referral opportunities.

17

17

Course Content (Philip's Part)

Date	Topics (Tentative)	Homework
December 5	Introduction to Causal Inference; Potential Outcomes Model	
December 12	A/B Testing; Regression; Matching	Problem Set 1
December 19	Instrumental Variables; Regression Discontinuity Design	Problem Set 2
January 2	Difference-in-Differences; Synthetic Controls	
January 9	Machine Learning Powered Causal Inference	Problem Set 3

18

18

Course Prerequisites

- No prerequisites, but some knowledge of statistics and coding will be useful.
- Highschool math is also assumed.
- Not adverse to programming (everything implemented in Python).
- Not adverse to analytical thinking and quantitative analysis in general.

Learning by Doing

- Lectures and demonstrations (recordings available on GitHub)
- Extensive cases and data: End-to-end analysis and problem-solving
- In-class discussions
- 5 problem sets and 2 projects (1 required and the other one as a bonus)
- Laptop
  - Bring a laptop to every class. Close your laptop until you are asked to use it.
  - Install the required applications (Python and Anaconda).
  - Download Jupyter Notebooks (from GitHub) to your laptop before each class.
- Attendance is required.

Course Materials

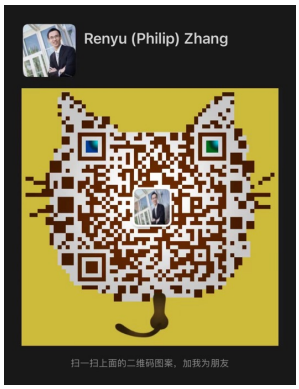
- Blackboard
  - Homework Submission
- GitHub:
  - https://github.com/DSME-6653/Web-W2023
- Anonymous Survey:
  - Link to the survey.
  - You are more than welcome to submit any feedback throughout this course.
- No required text books
  - Reference books given in the syllabus and provided at GitHub.

21

21

Course Communications

- Class Meeting: Tuesday, 6:30PM-9:15PM (@WMY_303)
- Office hour (Philip): Monday, 1:00PM-2:00PM, @CYT_911, or by appointment
- WeChat group: Online discussion forum.
- Instructor contact
  - Office: CYT_911
  - Email: philipzhang@cuhk.edu.hk
  - Tel: 852-3943-7763
  - WeChat: rphilip_zhang
- Teaching Assistant: Yilin Shi
  - Office hour: By appointment.
  - Email: ylshi@link.cuhk.edu.hk



22

22

Grading

- Class participation, 20%
  - Baseline 12%; lose 4% for each class missed; failing grades for missing 4 or more classes
- Problem sets, 30%=6%*5
  - **Due every one or two weeks after the distribution of a problem set**
  - 5 problem sets with the highest scores will count
- Projects, 20%
  - Project 1 (20%): Due on Tuesday, **February 20**
  - Project 2 (Bonus 5%): TBD
- Final Exam, 43%
  - **6:30PM-9:00PM, Tuesday, February 27**
  - Close-book, close-notes, electronic devices NOT allowed

23

23

Grading

- Problem sets and projects submitted via Blackboard:
  - Everyone should **individually** finish and submit his/her **own solutions**, though discussions with others are allowed.
- Regrading:
  - Submit your requests within **7 calendar days** after receiving your grade.
- "Zero-tolerance" policy
  - Any violation of academic integrity is strictly prohibited and will be treated seriously.

24

24

Who Should Quit This Course Now?

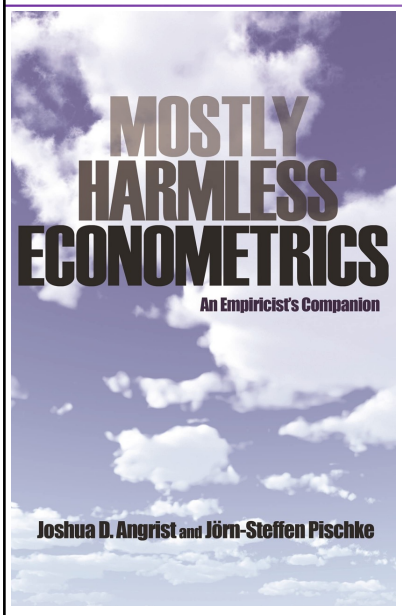
- You want an easy A.
- You hate the quantitative/analytical way of thinking and solving problems.
- You hate coding/programming.
- You hate me or Kevin.

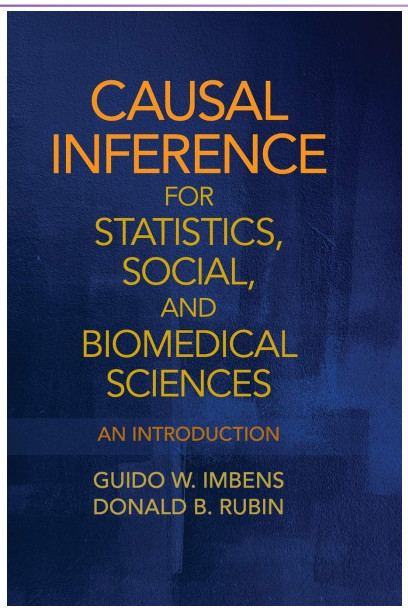
Otherwise, you are very much welcome joining us to enjoy the excitements and challenges of Web Analytics!

25

25

References





Angrist and Imbens' Nobel Prize Lecture

https://www.youtube.com/watch?v=DV4IjRfjTs0

Angrist's Lecture Series @ AEA

https://www.aeaweb.org/conference/cont-ed/2020-webcasts

An Online Tutorial with Python Code:

https://matheusfacure.github.io/python-causality-handbook/landing-page.html

Another Online Course:

https://www.bradyneal.com/causal-inference-course

26

26