



Algorithms, Big Data and Online Marketplaces

Welcome to the course!

Live Comments [弹幕]



Scan the above QR code using your WeChat.

Follow the Official Account and send your live comments.



Do You Know the Answer?

How many couples meet their partner online?

According to a Stanford University research project, by 2017, 39 percent of heterosexual couples reported meeting their partner online.

For homosexual couples, the rate is more than 70% now!



Here is the data

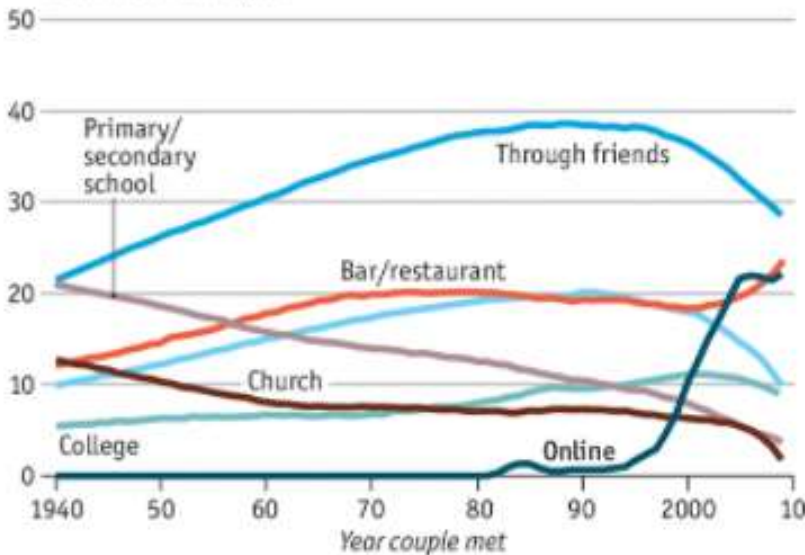
How couples met	1995, %	2017, %	Z score	Significance
Online	2	39	23.43	***
Through friends	33	20	-4.55	***
Through family	15	7	-8.47	***
Through or as coworkers	19	11	-5.16	***
In a bar or restaurant	19	27	2.38	*
In primary or secondary school	10	5	-6.62	***
In church	7	4	-2.52	*
Through or as neighbors	8	3	-4.54	***
In college	9	4	-1.17	

Here is the data (slightly outdated)

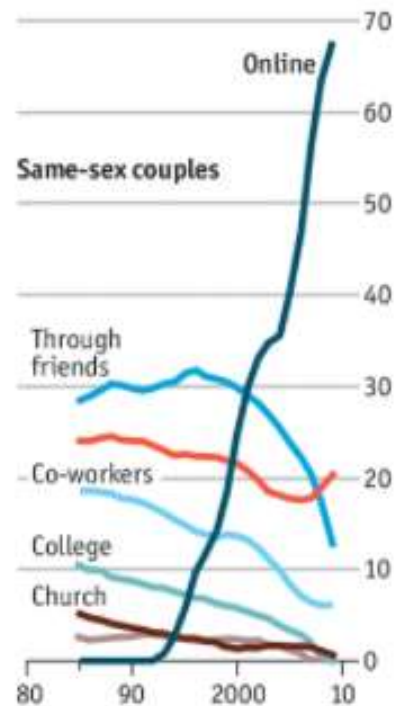
Meet markets

United States, how couples meet, %

Heterosexual couples



Same-sex couples





Internet creates wealth

Question: *What are the richest companies in the world?*





Internet creates wealth

Question: *What are the richest companies in the world in 2009?*





Internet creates wealth

Question: *What are the richest companies in the world in 1999?*



Largest Companies by Market Cap

In US Dollars

2012



Internet creates wealth

In 1999, the world's largest companies (by market capitalization) are

































Internet creates wealth

In 2009, the world's largest companies (by market capitalization) are



Internet creates wealth

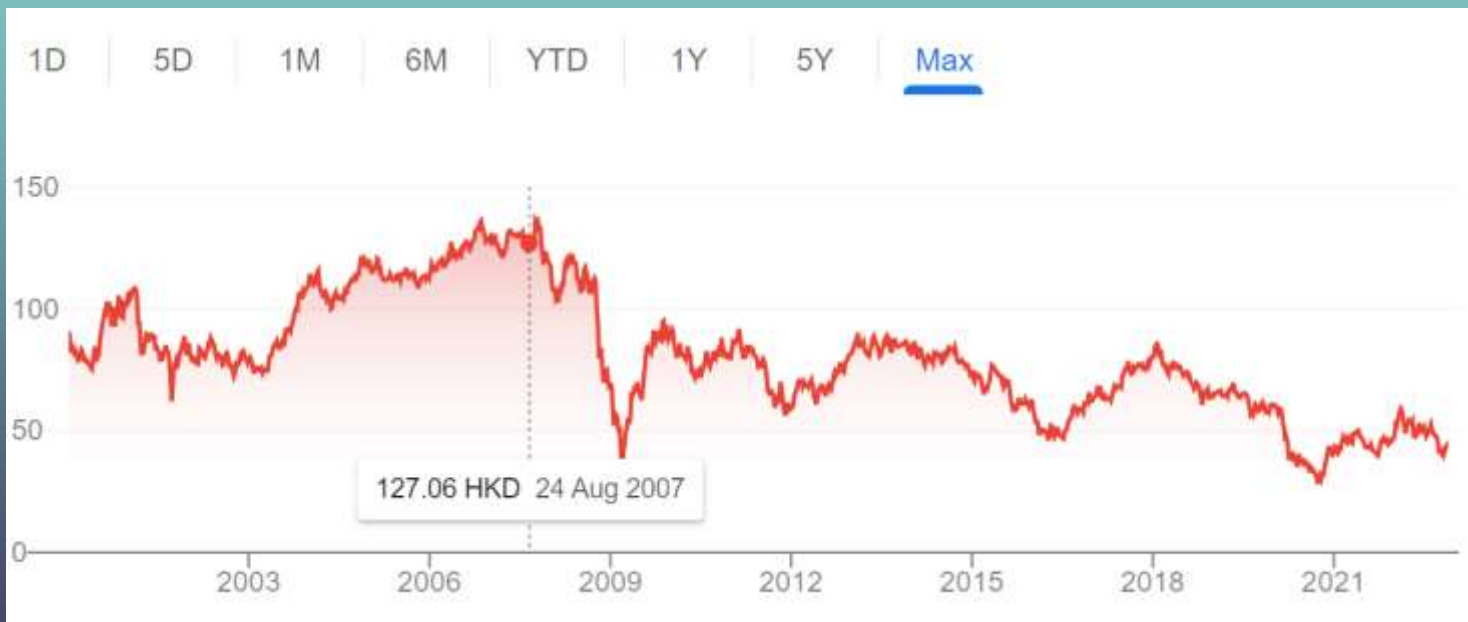
In 2023, they are

1	 Apple AAPL	\$2,908 T	\$184.92	-0.59%		 USA
2	 Microsoft MSFT	\$2,577 T	\$346.62	-0.43%		 USA
3	 Saudi Aramco 2222.SA	\$2,071 T	\$8.56	-0.62%		 S. Arabia
4	 Alphabet (Google) GOOG	\$1,571 T	\$124.06	-1.38%		 USA
5	 Amazon AMZN	\$1,287 T	\$125.49	-1.27%		 USA
6	 NVIDIA NVDA	\$1,054 T	\$426.92	-0.09%		 USA
7	 Tesla TSLA	\$825.78 B	\$260.54	-1.81%		 USA
8	 Berkshire Hathaway BRK.B	\$739.56 B	\$338.31	-0.44%		 USA
9	 Meta Platforms (Facebook) META	\$728.92 B	\$284.43	-0.92%		 USA
10	 TSMC TSM	\$542.34 B	\$104.57	-0.58%		 Taiwan

We know that HSBC is the largest bank in Hong Kong.



So, what's wrong with HSBC?



HSBC used to be a great investment!

圣诞钟买汇丰 [编辑]

文A 添加语言 ∨

条目 [讨论](#) [大陆简体](#) ∨

[阅读](#) [编辑](#) [查看历史](#) [工具](#) ∨

维基百科，自由的百科全书

圣诞钟，**买汇丰**，**香港股市**常见名词，其中“钟”和“丰”在**粤语**上是**押韵**，是指**传统**上在每年**圣诞节**前后（即12月尾至1月初）买入**汇丰控股**的**股票**，其股价或会有较好的**投资**回报，这情况在**牛市**上通常见效。美国股市同类理论叫**buy when it snows**。

Not only HSBC is in trouble...

Market Summary > Standard Chartered PLC

62.50 HKD

HKG: 2888

-9.74 (-13.48%) ↓ all time

+ Follow

17 Apr, 9:20 am HKT • Disclaimer

1D

5D

1M

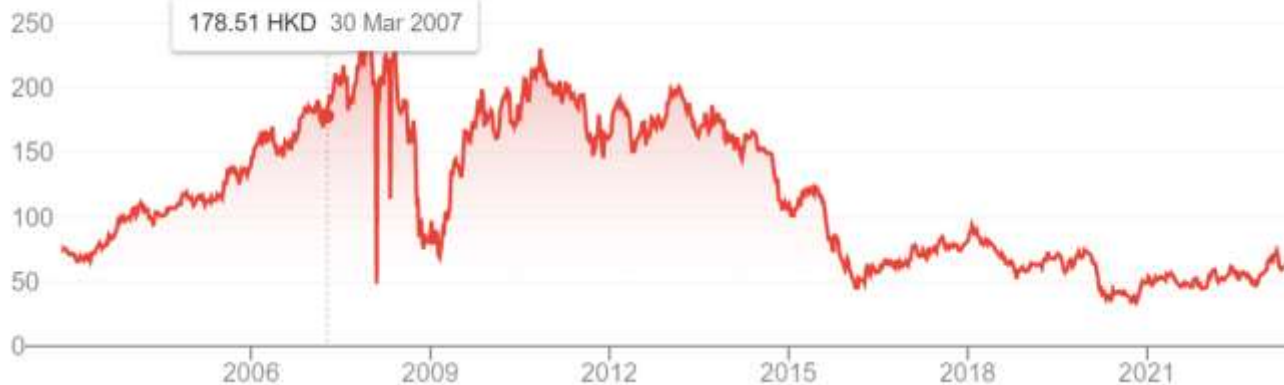
6M

YTD

1Y

5Y

Max



Not only HSBC is in trouble...

Market Summary > Citigroup Inc

49.56 USD

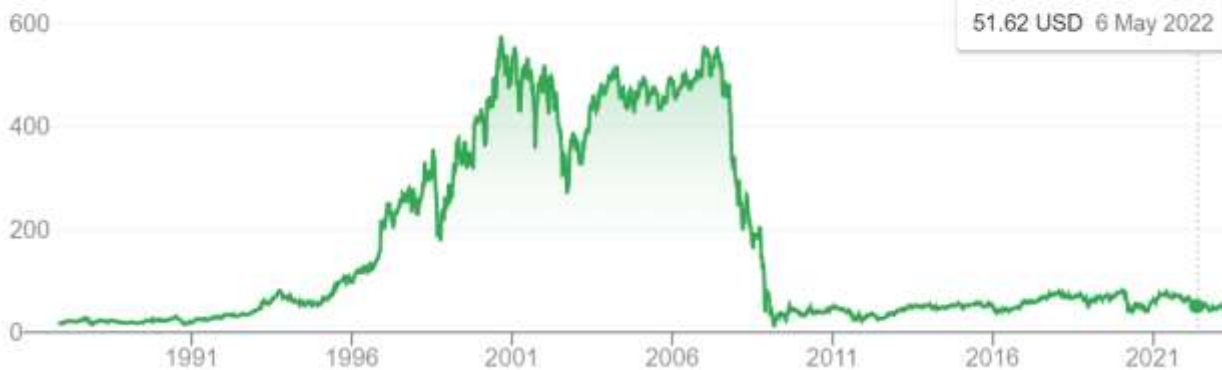
+ Follow

+30.91 (165.74%) ↑ all time

Closed: 14 Apr, 7:59 pm GMT-4 • Disclaimer

After hours 49.63 +0.070 (0.14%)

1D | 5D | 1M | 6M | YTD | 1Y | 5Y | Max



Not only HSBC is in trouble...

Market Summary > Bank of China Ord Shs H

3.11 HKD

-0.22 (-6.61%) ↓ all time

17 Apr, 9:20 am HKT • Disclaimer

+ Follow

1D

5D

1M

6M

YTD

1Y

5Y

Max



Reason: The rise of the Internet finance industry! But why?



What's this product?



It is a phone cradle!

But why do people purchase it?





Matthew Brennan @mbrennanchina · May 14, 2019

Chinese phone cradle for boosting your phone's daily step count. Some insurance companies in China allow people who consistently reach a certain daily step count to get discounted health insurance premiums.



Lê Nguyễn Hoàng (Science4All) @le_science4all · May 14, 2019

Replying to @mbrennanchina

Aka adversarial poisoning attack on data-driven algorithms...



9



97



Graeme Douglas ✓ @graemedouglas · May 14, 2019

Replying to @mbrennanchina and @BBHLabs

Not just China - Vitality does this in the U.K.



4



5



66



Uber collects data on your smartphone battery. But why?



Uber charges you higher prices when your battery is low,
because it knows you cannot afford to wait!






Uber charges you higher prices when your battery is low,
because it knows you cannot afford to wait!

Uber charges more from users if their phone battery is low, report claims

According to a small study by the Belgian newspaper Dernière Heure, Uber charged 6 per cent more for a journey made on a smartphone with only 12 per cent of battery remaining compared to the same journey.






Why do airlines raise prices if you keep searching on the same flight
(机票为什么越搜越贵)?


BUSINESS

The Truth About Whether Airlines Jack Up Prices If You Keep Searching the Same Flight




Many retailers offer free Wi-Fi to their shoppers. But why?






When you connect to a retailer's free Wi-Fi, the retailer will be able to monitor your entire browsing history: Did you search the same product online? What's your real time geographic location?





Taobao offers shopping cart, but Pinduoduo does not. Why?



Taobao usually offers users discount based on the value of products that you buy, but Pinduoduo offers users discount based on the number of users buying it.

“淘宝是引诱一个人买很多东西，而拼多多是引诱很多人买一个东西！”





Our class

The purpose of this class is very straightforward: we want to bring algorithms, big data and online marketplaces together to address the following questions:

How to collect data?

How to use analytics and algorithms to analyze data?

How to guide firms' business operations using the insights obtained from data?





Questions

What is an algorithm?

Do you know any algorithms?



Questions



ALGORITHMS

What is...

An algorithm?

Xi Li

What is an algorithm?

19/11/2023, 11:30:27



ChatGPT (Tokens used 97)

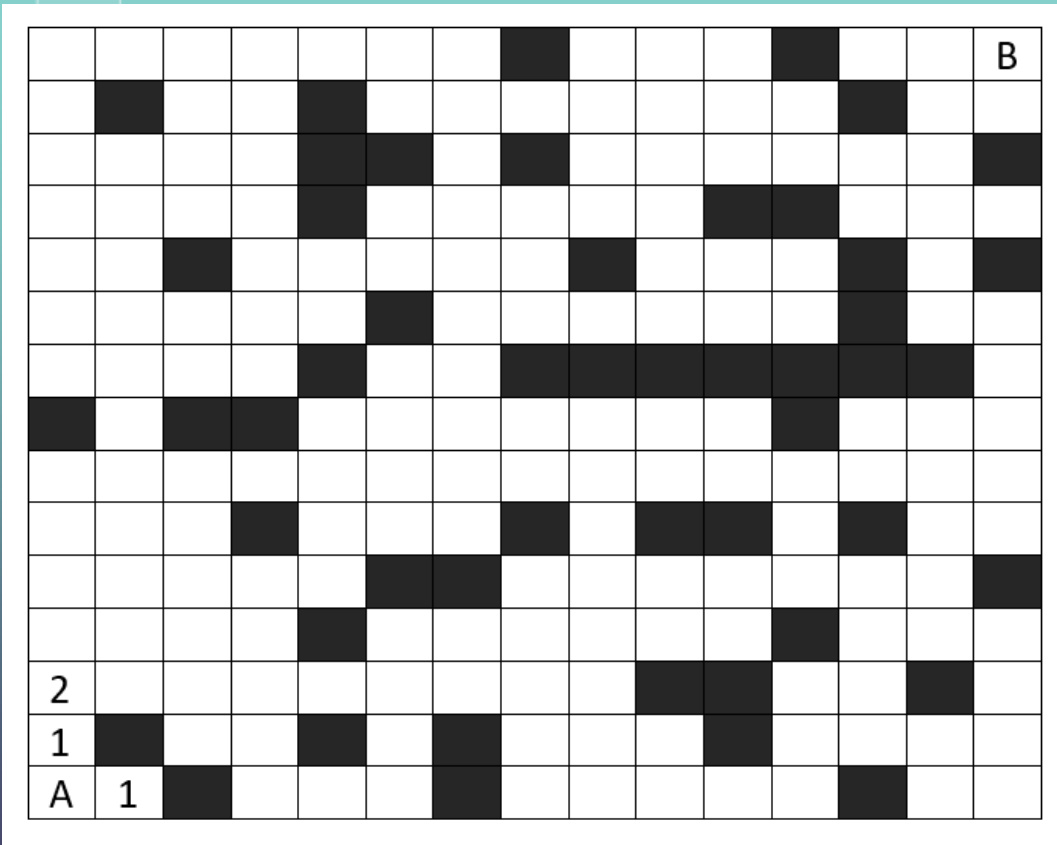
An algorithm is a set of instructions or steps that a computer program follows to solve a problem or perform a task. Algorithms are used in many areas of computing, including search engines, data analysis, and artificial intelligence. They are designed to be efficient and effective, and can be written in different programming languages depending on the needs of the application.

19/11/2023, 11:30:27

This is perhaps the only B-school class with an “algorithm” in its title!

							■				■			B
	■			■								■		
				■	■		■						■	
				■					■	■				
		■					■				■		■	
				■							■			
				■			■	■	■	■	■	■		
■		■	■								■			
			■			■		■	■	■		■		
				■	■	■							■	
				■					■	■			■	
1	■			■		■				■				
A	1	■				■						■		

These cells can be reached in 1 step!



These cells can be reached in 2 steps!


16	17	16	17	18	19	20		22	23	24		28	27	B
15		15	16		20	19	20	21	22	23	24		26	27
14	13	14	15			18		20	21	22	23	24	25	
13	12	13	14		16	17	18	19	20			25	26	27
12	11		13	14	15	16	17		19	20	21		25	
11	10	11	12	13		15	16	17	18	19	20		24	23
10	9	10	11		13	14								22
	8			11	12	13	14	15	16	17		19	20	21
6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
5	6	7		9	10	11		13			16		18	19
4	5	6	7	8			11	12	13	14	15	16	17	
3	4	5	6		8	9	10	11	12	13		17	18	19
2	3	4	5	6	7	8	9	10			17	18		20
1		5	6		8		10	11	12		16	17	18	19
A	1		7	8	9		11	12	13	14	15		19	20

Repeat this again and again until you can reach cell B!



Examples of Algorithms

The Netflix Prize: In 2006, Netflix launched a competition called the Netflix Prize, offering a \$1 million prize to anyone who could improve their movie recommendation algorithm by 10%. This attracted thousands of participants from around the world, including data scientists, mathematicians, and computer scientists. It took several years, but eventually, a team named “BellKor's Pragmatic Chaos” won the prize in 2009, improving Netflix's recommendation algorithm by 10.06%





NETFLIX 2009
DATE: 09.21.09
PAY TO THE ORDER OF: BellKor's Pragmatic Chaos \$1,000,000⁰⁰
AMOUNT: ONE MILLION 00/100
FOR: The Netflix Prize Reed Hastings

The Netflix Prize



Who should take this course?

MSc Students who are

- Interested in marketing in the digital age.

- Interested in data analysis and programming in general.


- Interested in joining a big-tech firm or start a career in the Internet industry.





Social Listening

Through the years, colleagues, students, and TAs share students' online chatter with me. Now, let me share some with you.



mkt本学年第一门纯quant课程结课啦

7025拥有最长的课程名字

Algorithm, Big Data, and Online Marketplace

一门看上去蛮无聊但实际很好玩的课

.

📖 课程

从各种regression到causality

从R到Tableau到webscraping

在marketing的课程里确实算很硬的了

但老师从基础开始手把手包教包会!

coding零基础和文科生也完全不用担心

课上匿名实时弹幕好用又好笑

方便大家肆无忌惮的提问(和吐槽)

.

🎤 老师

清华大牛 编程能手 天生喜剧人

不用过节工作狂 mkt泥塑艺术家

高频使用短句：又不难

While the comments are inaccurate...
I am not a workaholic or an artist...
But you are encouraged to post live
comments!

2 课程安利

👉 7025 算法/大数据技术的营销应用（选修）

这门课的英文名是Algorithm, big data and online marketplace ~ 这门课相当硬核了！会涉及编程、数据挖掘、数据可视化处理和机器学习的管理方面的知识

🤖️但没有编程基础的小白完全不用担心和害怕！和蔼可亲的教授会在课上亲自示范编程的每一个步骤，而且Xi Li教授很可爱，课上我们可以在他自制的弹幕页面发弹幕互动！很有趣对不对 🥰 🥰 🥰

是会用望远镜和实时弹幕的可爱港大教授!

只能说Xili老师长得就很清华😏

7025的课超级有用，workload也正正好好（coding入门）

最后还能带你感受直男的浪漫（他会送每个人一份课堂定制小礼物📺）

虽然老师的英语会有点口音，但非常自信流利，习惯就好！！

#日常碎片PLOG #笔记灵感 #港大商学院 #香港大學 #香港大学 #港大mkt

Xi Li: I don't know what to say...

💖 7025 Algorithms, Big Data and Online Marketplaces

这门课我直接吹爆！ Xi Li是我在港大最喜欢的老师 🏠

（清华计算机本科的大佬）他的课上可以发弹幕随时提问（中文也完全ok），而且课程有趣+干货满满！我在这门课超积极哈哈 🍓 课程主要结合case学习数据分析的方法、原理和R语言实操，包括线性回归、逻辑回归、固定效应、二次回归、敏感性分析，还介绍了网页爬虫、tableau和前沿研究。Workload较大但收获颇丰（btw Xi的Chinglish听起来超亲切 🍓）

Workload is not heavy... Hope my Chinglish has improved...

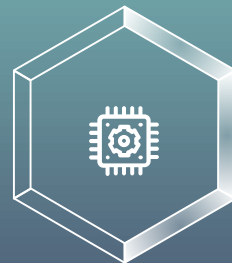
This course adopts methodologies from



**Computer
Science**



Statistics



Economics




Leveraging your competitive advantage

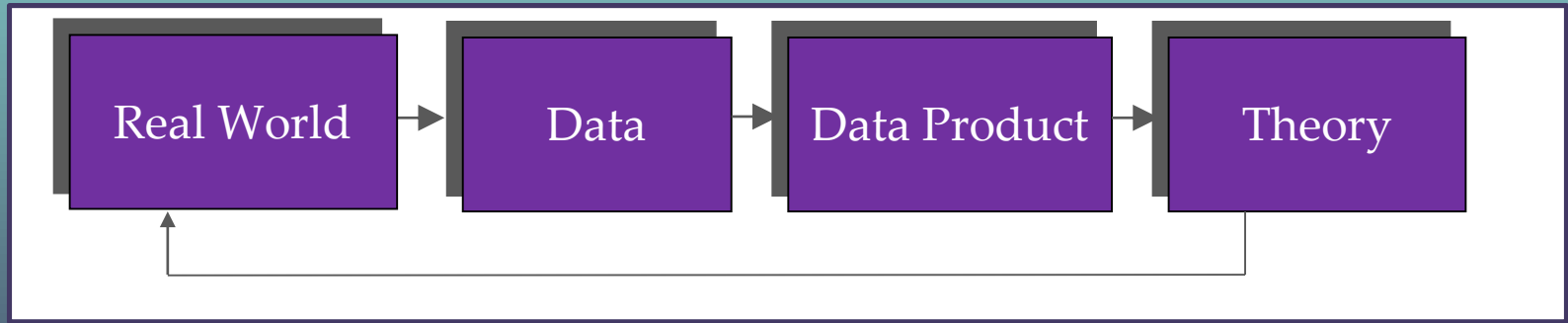
Compared to traditional marketers, you know how to program and how to analyze data.

Compared to statisticians and computer scientists, you understand consumers better, and know how to apply results to business settings.

Compared to economists, you not only know the theory, but also know how to apply the theory to solve real-world marketing problems, and test the theory using real-world data.



The LOOP





It seems that our class motto is

“Talk is cheap, show me the data!”





But actually, it is

If you torture the data long enough, it will confess to anything.

如果你折磨数据足够长的时间，它们会招供的。

---Ronald Coase (Nobel Prize Winner)



We emphasize on data, data, and data

Recall that the program offers you an **MSc** degree in Marketing, not an **MA** degree.

Why? Because you are learning quantitative methods, not just how to make presentations and talk about business ideas.



HKU
BUSINESS
SCHOOL
港大經管學院

Master of Science in Marketing
理科碩士(市場營銷學)

We solve problems for



Firms



Platforms



Individuals

Do I really want to take the course?

I don't know. It depends. But let me offer you some advice.

Take the course if

You are interested in data analysis, and you are considering to become to data scientist in the future.

You want to understand how the online marketplaces work.

You want to explore rigorous research methodologies from different areas.

Don't take the course if

You hate data analysis or programming.

You already know the materials to be covered in the class.

You want to pick an easy course to fulfill your credit requirement.



Additional Course Information



The Instructor

Xi Li, Professor of Marketing
Director, Asia Case Research Centre
Associate Director, Institute for Digital Economy and Innovation

PhD in Management, University of Toronto.
M.Phil. in Operations Research, HKUST.
B.E. in Computer Science, Tsinghua University.

Research interests: Algorithms, big data and online marketplaces.
Email: xili@hku.hk
Office Hour: by appointment



Our Teaching Assistant

Ray Song

Can be reached via email: soongray@hku.hk



My Research

We observed an interesting phenomenon: In a crowdfunding campaign, two very similar products often have very different prices. See the following examples. We addressed this puzzle.

LIFE ON THE LINE



A unique book of portraits connecting and celebrating life along the Arctic Circle.

Created by:
Cristian Barnett

201 backers pledged £12,134 to help bring this project to life.

Last updated September 28, 2014

£30

THE BOOK. A signed copy of LIFE ON THE LINE, your name included as a supporter in the book and an invitation to the launch.

£150

THE SPECIAL BOOK. A collector's edition of LIFE ON THE LINE (limited to just sixty-six copies), real cloth and slipcase, signed with a personal dedication by Cristian Barnett, invitations to events, plus special thanks in the book and online.

My Research



We are familiar with HK's local online marketplace, HKTVMall. Interestingly, at HKTVMall, consumers can observe the historical prices of products that they are interested in, a feature not available at other major online sellers.

We ask: Why is HKTVMall willing to disclose the historical prices? What happens when consumers can observe it?

My Research



Back to this question.

We investigate what happens when consumers can manipulate data. How should firms manage consumer manipulation? Does data manipulation always benefit consumers? Should regulators be concerned with consumers' data manipulation?



Course Website

In addition to the official course website provided by the university, we will also use a semi-official course website. You can find the latest materials and updates on this course website.

https://ximarketing.github.io/_pages/teaching/

Password for ABOM: 7025

Please do not share the course content with others.





Textbook? No.
Real Data? Yes.






Course Overview

Data analytics with R.

You can use Excel or SPSS, but they are too simple and cannot handle complex data analysis projects.

R is a free software that is commonly used for statistical analysis.

It is not only useful for digital marketing, but also useful for other purposes such as machine learning and optimization.





Course Overview

Data visualization with Tableau.

Data visualization gives us a clear idea of what the information means by giving it visual context through maps or graphs.

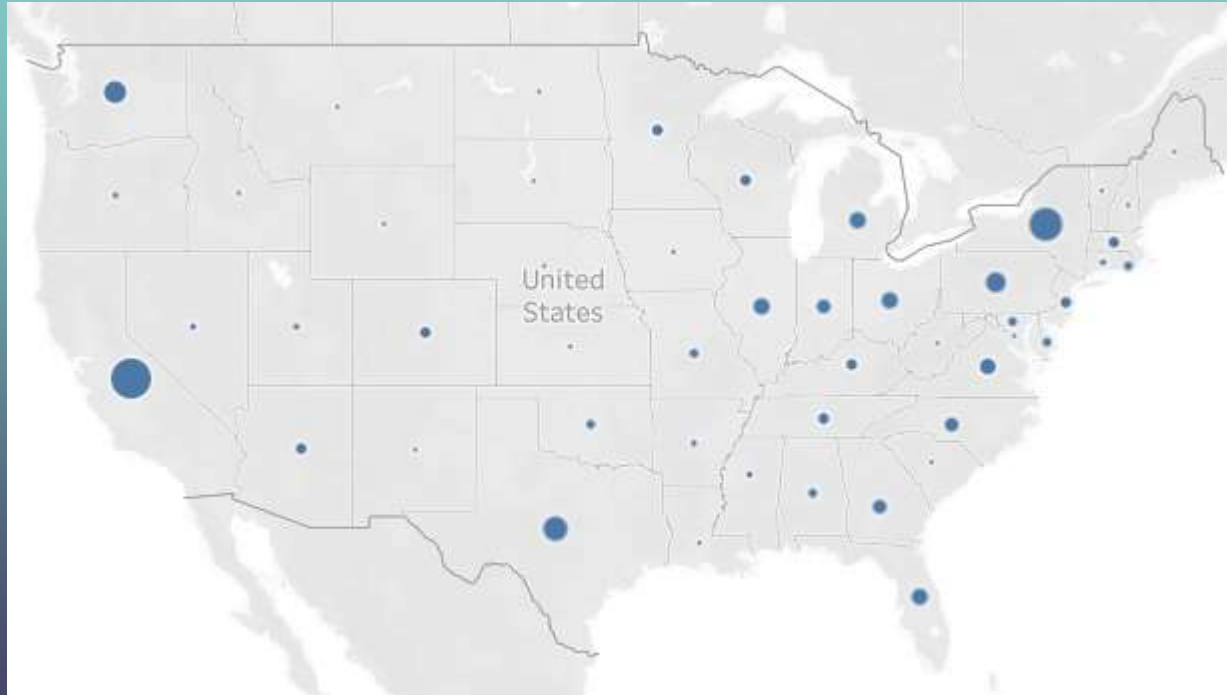
When you want to show your results to others, the best way to use figures --- *“A picture is worth a thousand words”*.



Tableau

United States California	United States Texas	United States Illinois	United States Ohio	United States Michigan	United States Virginia
	United States Washington	United States North Carolina	United		
		United States Indiana	United States		
United States New York	United States Pennsylvania	United States Georgia			
		United States			
	United States Florida	United States			

Tableau



Course Overview

Web Scraping with R.

Data scraping allow you to scrawl information from websites, e.g., online marketplaces. For example, you can collect product information from HKTV Mall, weather information from Hong Kong observatory, and tweets from Twitter.

Warning: Do not expect that you can become an expert in data scraping within one class. This is simply an introduction, and you need more practice yourself.




Course Overview

Web Scraping with R.

Why choosing R? R is powerful and is widely used for data analysis.

Python is also a great choice (and maybe a better choice for machine learning tasks). However, it is not that friendly for beginners --- you need to take a more specialized course to learn it.



Course Overview

This may be the only web scraping class in a business school (I don't know the answer, though).

We will see how to gather information from HKU faculty webpage:






Course Overview

This may be the only web scraping class in a business school (I don't know the answer, though).

Note: We are only going to cover the simplest techniques for web scraping. If you want to learn something more complex (e.g., how to login to your Moodle and download your materials automatically), you still need take some additional courses.



Course Overview

2021 Nobel Prize in Economics



Joshua D. Angrist and Guido W. Imbens

“for their methodological contributions to the analysis of causal relationships”

Course Overview

Causality

What is the fundamental difference between economics/marketing and statistics/machine learning? It is causality.

In statistics and machine learning, we ask if X **predicts** Y. But now, our question is, does X **cause** Y?

We will talk about when and how to draw conclusions on causality.



Course Overview

Logistic Regression

You should be already familiar with linear regression, the simplest statistical model for predicting.

But linear regression only works for certain dependent variables, and it works poorly with binary dependent variables.

Logistic regression is introduced to deal with the issue.



Course Overview

Going beyond logistic regression

Logistic regression bears some similarities with some human tasks such as autonomous driving and digit recognition.

Indeed, our human brain also calculates logistic functions.

A fundamental machine learning algorithm, artificial neural network, is a generalization of the logistic regression we discussed.

Course Overview

Text Analysis

In the past, we focus merely on numerical data (e.g., sales, profit, purchases, price, time etc.).

However, today, most of the data take other forms. Many of them are text information.

This includes online reviews, product descriptions, Tweets, SMS messages, forum discussions, firm announcements etc.



Course Overview

Text Analysis

We are going to take some simple measures to extract meaningful information from text data.

Sentiment analysis: It classifies text based on sentiment polarization (positive vs. negative).

Latent Dirichlet Allocation (LDA): It classifies text based on the topic of the text.



Course Overview

Price Personalization

Do you know behavior-based pricing (杀熟)? Why does firm charge higher prices to old consumers and lower prices to new consumers? Is doing so profitable?

More generally, how should firms use consumer data to offer them personalized prices to improve profit?

Should public policymakers regulate price discrimination based on big-data technologies?

Course Overview

Price Personalization: Policy considerations

中华人民共和国个人信息保护法

(2021年8月20日第十三届全国人民代表大会常务委员会第三十次会议通过)

第二十四条 个人信息处理者利用个人信息进行自动化决策，应当保证决策的透明度和结果公平、公正，不得对个人在交易价格等交易条件上实行不合理的差别待遇。

通过自动化决策方式向个人进行信息推送、商业营销，应当同时提供不针对其个人特征的选项，或者向个人提供便捷的拒绝方式。

通过自动化决策方式作出对个人权益有重大影响的决定，个人有权要求个人信息处理者予以说明，并有权拒绝个人信息处理者仅通过自动化决策的方式作出决定。



Course Overview

Recommender Systems

Every time you visit Amazon, Taobao and YouTube, you always receive some recommendations from these platforms.

The recommendations are made based on your past behavior and characteristics of the products/services.

We will talk about how online platforms make personalized recommendations using big-data technologies.



Grading

50% Group work:

40%: Two data projects, 20% each.

10%: Providing an interesting observation from online (or even traditional) marketplaces (possibly with your own explanations; this is *optional*), and present it in class. You also need to join the discussion of other groups.

20% In-class participation:

Class attendance and participation in discussions. (TA will take notes).

30% Individual Assignment:

Data Collection and Analysis




Data Projects

In this course, we are introducing one (or two?) practice class on data analysis. This is one unique feature of our course.

You have already learned about R and Tableau. Perhaps you can also use many other tools such as SPSS, Excel, Python etc.

I will give you some real-world business data, and your goal is to study the data using the tools you know --- It's entirely up to you how you want to play with the data!






Data Projects

You are going to work with your teammates on the data projects. Then, each group will submit a report illustrating the findings from the data. **It will be graded, and it affects your final grade.**

We will start with the data analysis in the class, and you will finish all the analysis after the class.


The purpose of the data project is to help you understand the data analysis methodologies and gain a sense of real data that data scientists are analyzing everyday.





Overall Course Structure

We have 10 lectures over the entire module. Among them, we are going to have

- 1 introduction class (today)
 - ~1.5 data workshops
 - ~7.5 regular lectures
- 



Presentation

In the class, each group is required to provide an interesting observation from online (or even traditional) marketplaces, and invite other groups to discuss about it. You may also share your own thoughts and discuss about other groups' topics. This counts for 10 points toward your final grades.

I will give you examples later on.


The grading is based on the “interestingness” of your observation (5 points), as well as your discussion of other groups' topics (5 points).





Presentation

The presentations/discussions will be in the beginning of a few lectures, and the order of presentation will be drawn randomly after group formation is completed.





TODO List

If you decide to take this course, here is what you should do after today's class (Important!):

Form groups with your classmates.

Individual task: Install R on your laptop. We will be using it in the next class.

Bring your laptop with you for the next lecture!



Group Formation

Each group consists of 6~8 students.

You need to **choose a name** for your group, e.g., “Marketers”, “Fantastic”, “A Plus” ...

The deadline for submitting your group information is:
23:59 Dec 12 (Tuesday; one day after the add-drop period)

Let the TA know if you cannot find a group.



Let's Download and install R.

Your installation path must not contain any non-English characters. Otherwise, you will have troubles using it.

安装路径必须为纯英文，否则运行可能出错。





Next, let's download R-Studio.
It is also free.

Your installation path must not contain any non-English characters. Otherwise, you will have troubles using it.

安装路径必须为纯英文，否则运行可能出错。

