

# Algorithms, Big Data, and Online Marketplaces

Scan the QR code using WeChat to post live comments on our screen:



What are the wealthiest firms in the world?

What were the wealthiest firms in the world in 2010?

<https://www.youtube.com/embed/Z93yWXb9Tb0?enablejsapi=1>

HSBC is the largest bank in Hong Kong.  
Let's see HSBC's stock performance.

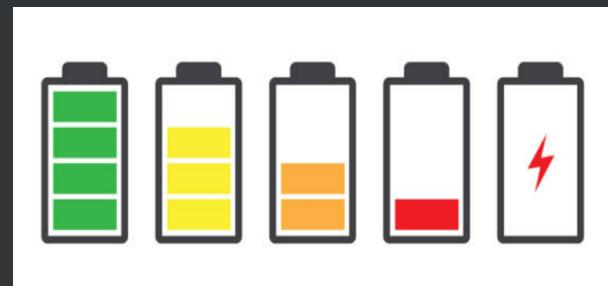
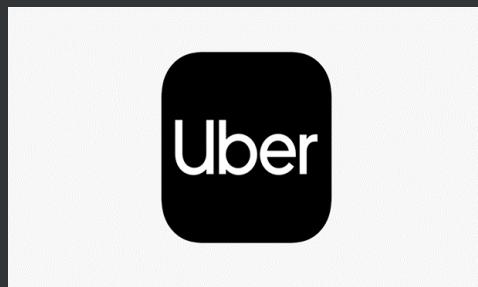
## 何謂「聖誕鐘買滙豐」？

- ◆ 這句話是早年本港股壇流行的傳統智慧
- ◆ 意指在12月這聖誕月買入滙豐股票，一般都能獲利
- ◆ 這句話的來由，坊間有不同說法



What's wrong with HSBC?

Uber, the leading ride-hailing APP, collects data on your smartphone's battery level. But why?



Uber charges you higher prices when your battery is running low.

## **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.

More specifically, when your battery is less than 12%, Uber will charge you 6% more for your journey.

Lenddo, a Singaporean based start-up, helps financial institutions collect your social network data, but why?



The firm analyzes your social network to estimate your credit score and decides whether to offer you a loan.

**MARKETING SCIENCE**

*Articles in Advance*, pp. 1–25

ISSN 0732-2399 (print) | ISSN 1526-548X (online)



<http://dx.doi.org/10.1287/mksc.2015.0949>

© 2015 INFORMS

## Credit Scoring with Social Network Data

Yanhao Wei

Department of Economics, University of Pennsylvania, Philadelphia, Pennsylvania 19104, [yanhao@sas.upenn.edu](mailto:yanhao@sas.upenn.edu)

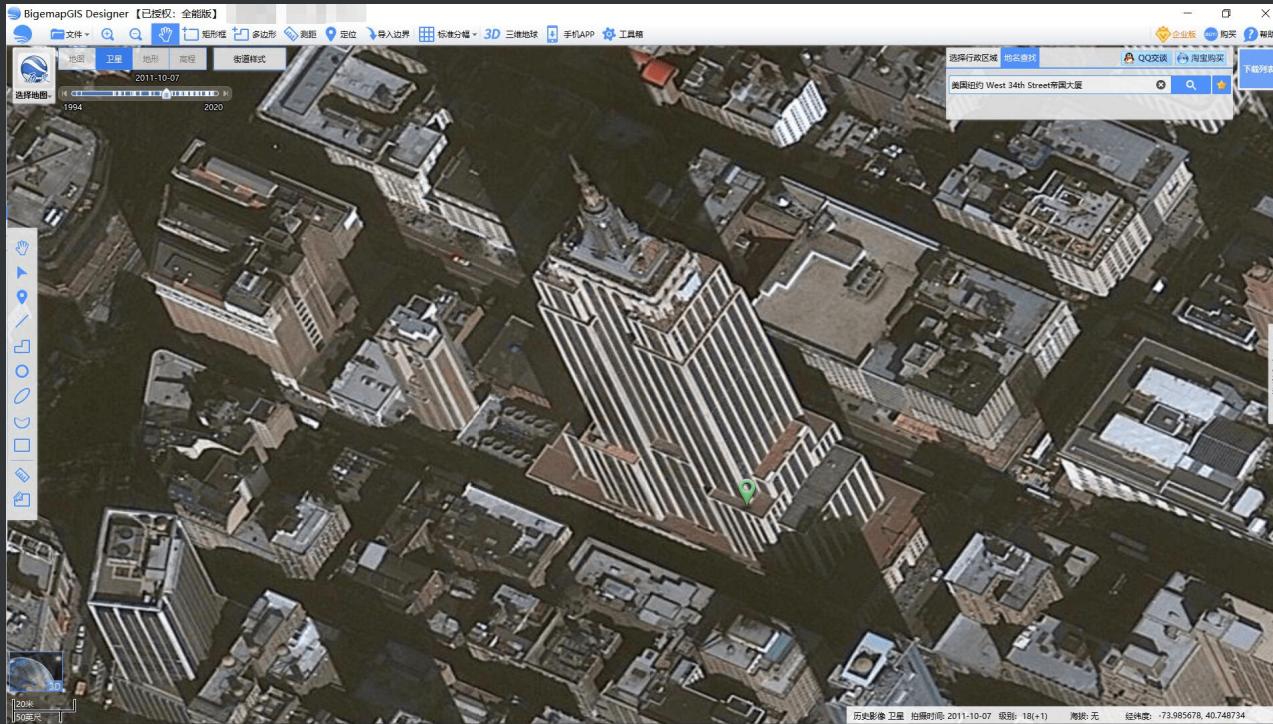
Pinar Yildirim, Christophe Van den Bulte

Marketing Department, The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104  
[pyild@wharton.upenn.edu](mailto:pyild@wharton.upenn.edu), [vdbulte@wharton.upenn.edu](mailto:vdbulte@wharton.upenn.edu)

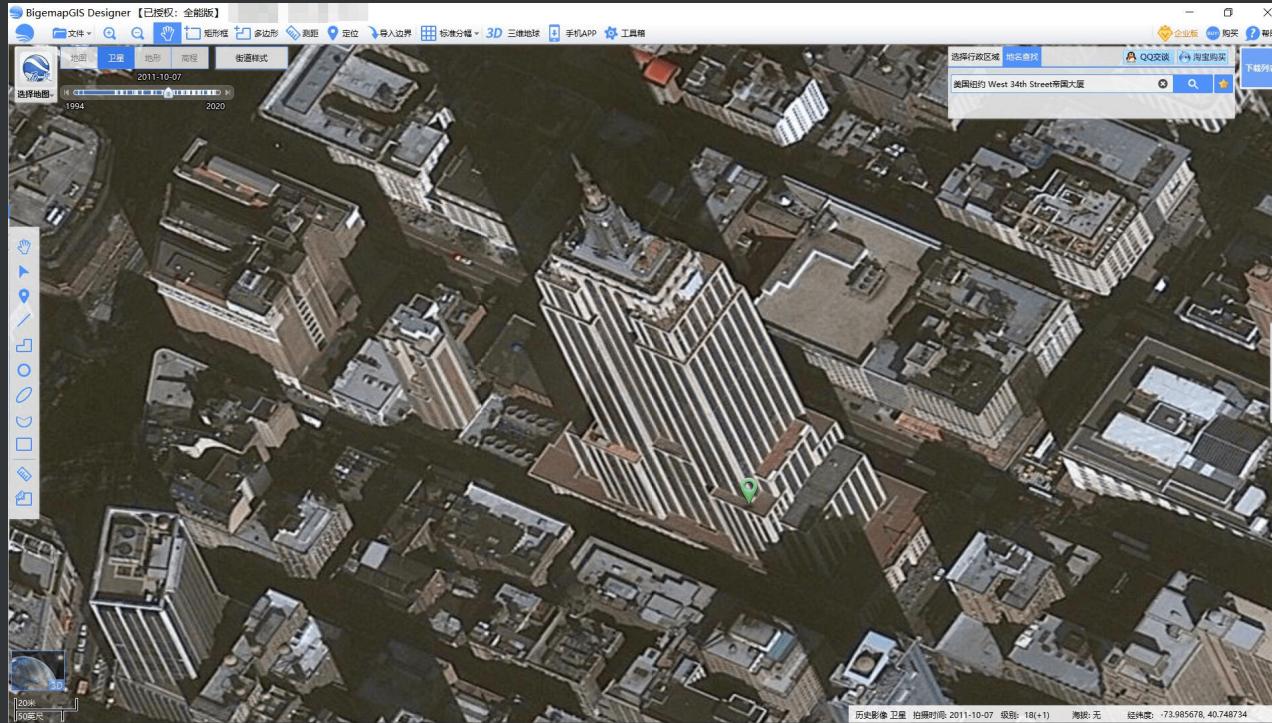
Chrysanthos Dellarocas

Information Systems Department, Questrom School of Business, Boston University, Boston, Massachusetts 02215, [dell@bu.edu](mailto:dell@bu.edu)

US retailers Target and Kroger both purchase real-time satellite images from RS Metrics. But why?



They monitor the competitors' carpark to estimate the competitors' performance and revenue.



What is this?



It is a phone cradle.  
Why do people spend money buying 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 Nguyên Hoang (Science4All)** @le\_science4all · May 14, 2019

Replying to @mbrennanchina

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



1 ↗ 9

1 ❤ 97



**Graeme Douglas** ✅ @graemedouglas · May 14, 2019

Replying to @mbrennanchina and @BBHLabs

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

1 4

1 ↗ 5

1 ❤ 66



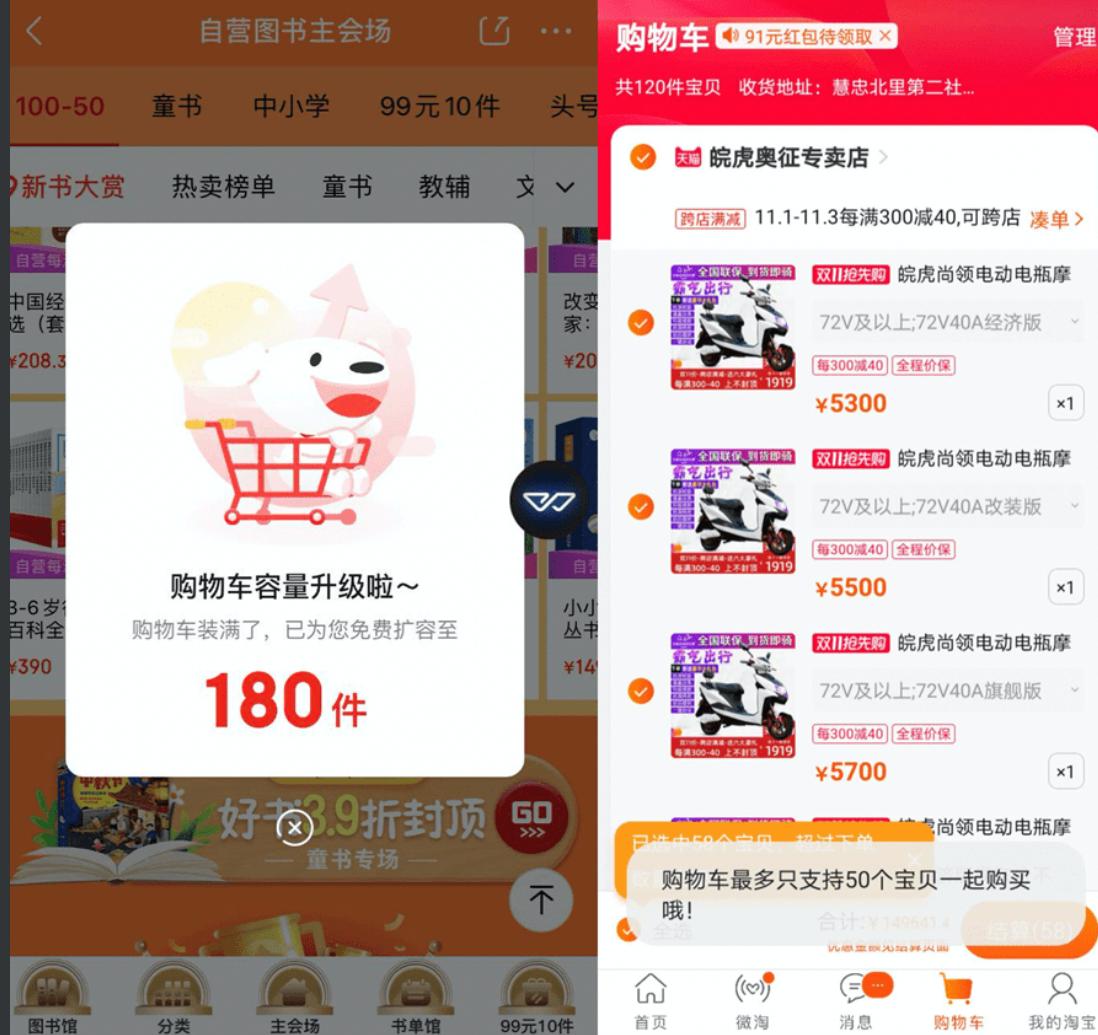


# Netflix: How did it know I was bi before I did?

After BBC reporter Ellie House came out as bisexual, she realised that Netflix already seemed to know. How did that happen?

Why do offline sellers offer free Wi-Fi to users?





Why does Taobao offer a shopping cart, but PDD does not?

Researchers study the employees of a big firm, and find that these employees tend to spend more money during the 11th day of every month. What is the reason?

11th is the payday of the company!

Online sellers are actively collecting data from your banking APPs (this becomes illegal after the passing of Personal Information Protection Law in 2021). They can then learn about your payday and send you coupons and deals on or before your payday.

# Applications of Big Data

## Demand Forecasting



Cinelytic is a Los Angeles based start-up. It collects data on title name, synopsis, logline, genres, rating, per-territory distributors, production budget, keywords, associated production companies, per-territory theatrical revenues, social media metrics, and other relevant data to predict the box office revenue of a new movie.

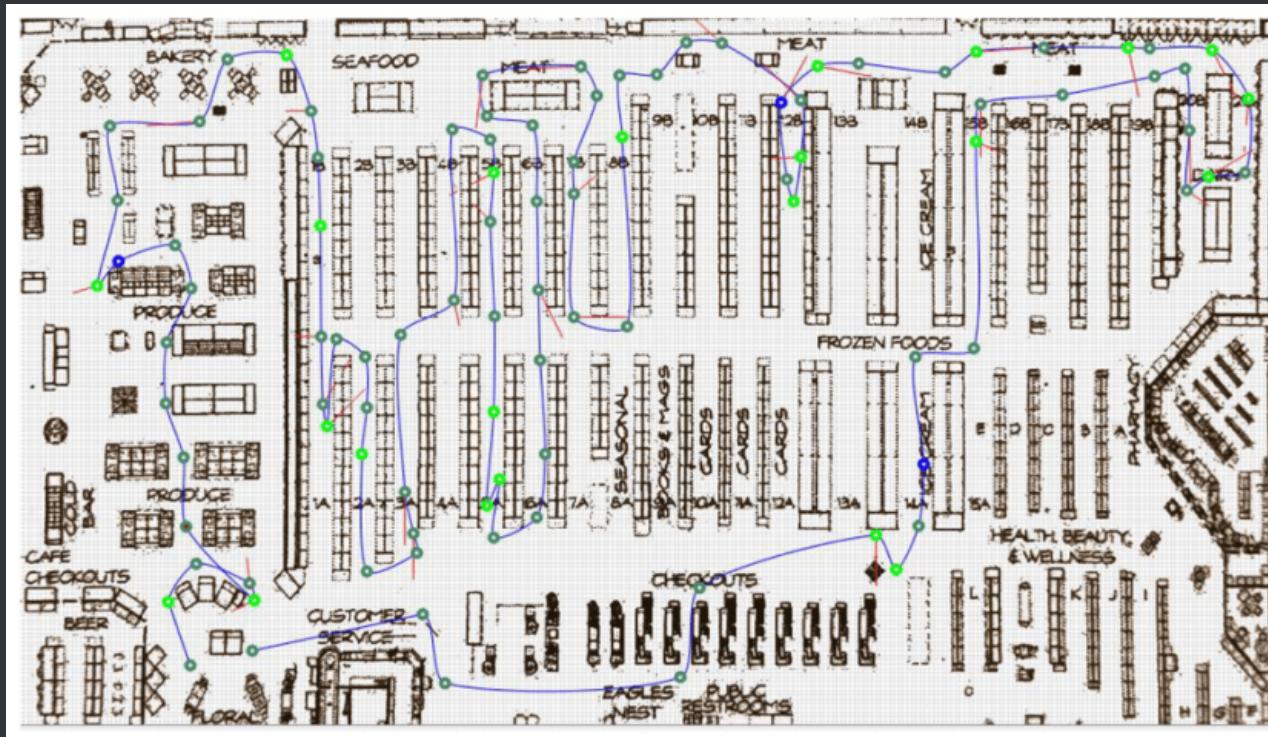
## Product Management

Prof. [Raymond Burke](#) of Indiana University was hired by a chain supermarket in the US, where he manages a number of shelves. He installed video cameras in store and invited users to wear eye-tracking system to analyze store flow and eye movement. Based on the data, he optimized the layout of the store and product, improving store revenue by 30%. You can find his video [here](#).

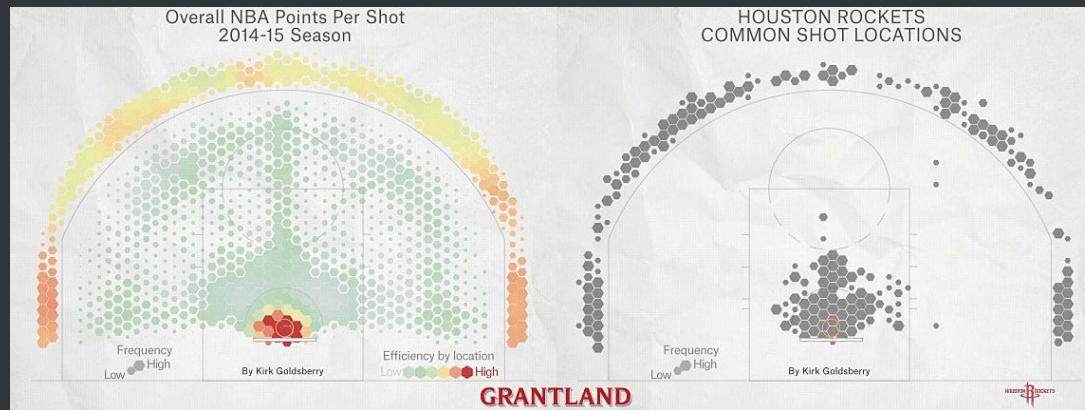
# Product Management



# Product Management



# Opponent Analysis



Every NBA team employs data scientists who analyze their own strengths and weaknesses, as well as those of their opponents. These professionals collect various types of data, including information on ball and player location, speed and distance, touches, passing, shots, and behavioral habits, among others.

Data Analytics

# **Data champions: How the Golden State Warriors are turning on-court data into a competitive advantage**

March 5, 2021

## Consumer Analytics



Nemesysco is an Israeli start-up. It helps call centers monitor phone calls to detect emergences and irregularities. It also helps insurance companies detect fraud through analyzing phone calls.

## Personalized Recommendation

有人只是去参加了一次北京婚博会，晚上回到家打开微博和微信，发现信息流广告全部变成了婚纱照、婚庆公司、婚礼礼服等。令他感到恐怖的是在此之前从未在手机进行过结婚相关的任何搜索。这一切发生改变的原因仅仅是因为他去了一次婚博会这个地方而已。

有人在知乎看到除甲醛的相关问题，只是百度了一下，结果连一个美食app都开始推荐除甲醛公司。在百度上打开某理财网站，不到半小时推销电话就打过来了。

有时候你在网上搜了一本小说，然后突然就会有很多假网站在百度上显示他们网站有这本小说可以下载，然后让百度把他推在首页，你打开链接一看其实里面没有，但是有其他东西的广告。

## Personalized Pricing

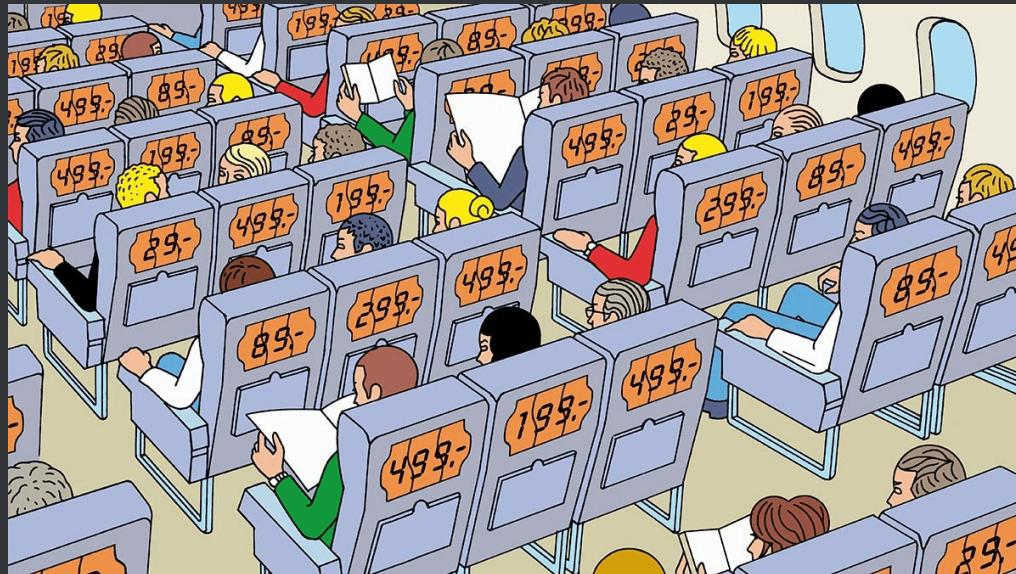
The screenshot shows the homepage of The Wall Street Journal. The main title "THE WALL STREET JOURNAL." is prominently displayed in large, bold, black letters. Below it is a navigation bar with links to "English Edition ▾", "Print Edition", "Video", "Podcasts", and "Latest Headlines". A secondary navigation bar below includes "Home", "World", "U.S.", "Politics", "Economy", "Business", "Tech", "Markets", "Opinion", "Life & Arts", "Real Estate", and "WSJ. M". The main headline, "On Orbitz, Mac Users Steered to Pricier Hotels", is displayed in large, bold, dark blue letters. The text of the article begins below the headline.

# On Orbitz, Mac Users Steered to Pricier Hotels

The online traveling agency Orbitz charges users accordingly to their device. iOS users pay 30% more than Windows users do.

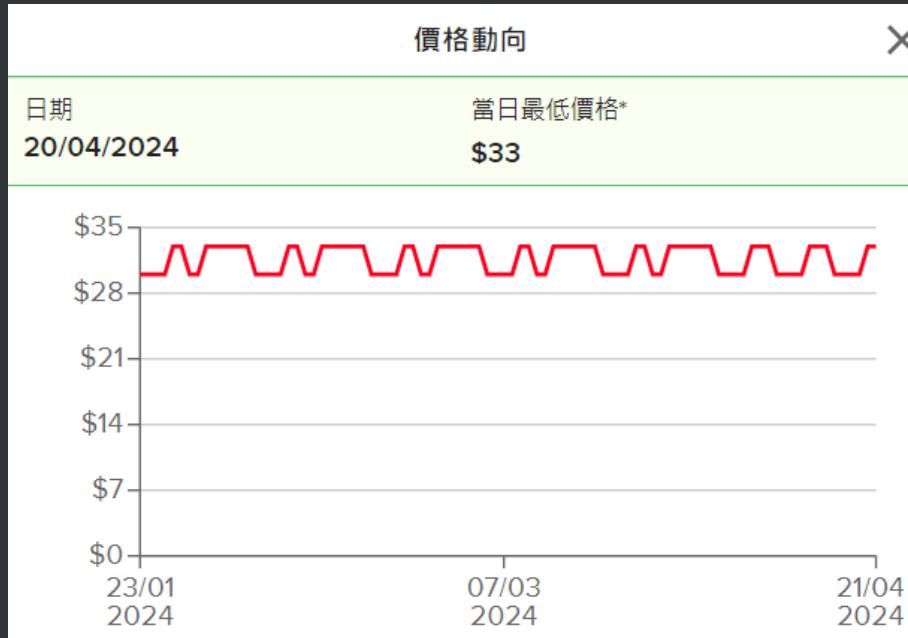
The online traveling agency Orbitz charges users accordingly to their device. iOS users pay 30% more than Windows users do.

## Personalized Pricing



Airlines and OTAs have long been known as varying prices according to your geographic locations (IP addresses).

## Traditional Pricing



Platforms such as Amazon automates their prices even every minute.

# What's an algorithm?

You see, we have “algorithms” in the course title, but what  
an algorithm is?

# Don't ask me. Ask your GPT.

 Xi Li

What's an algorithm

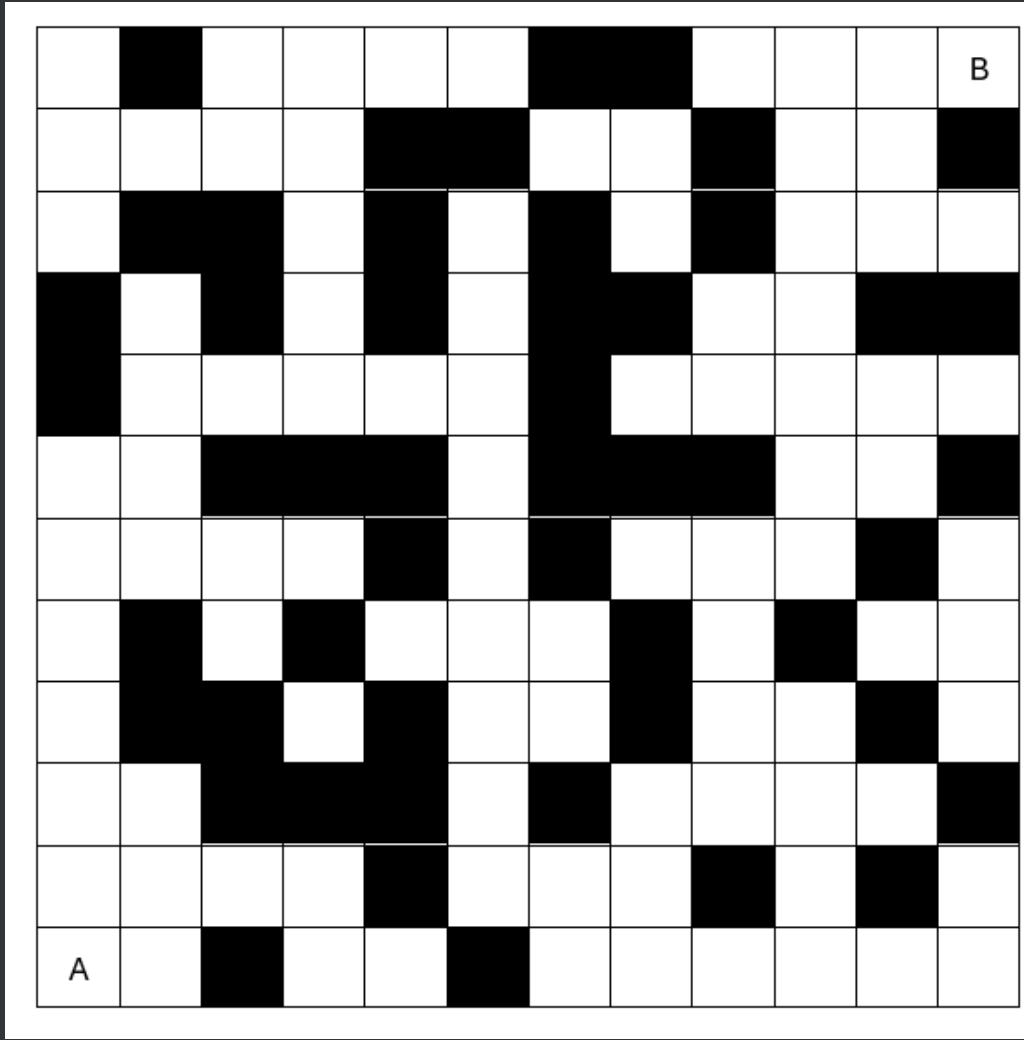
25/04/2024, 08:09:11

 ChatGPT (Tokens used 109)

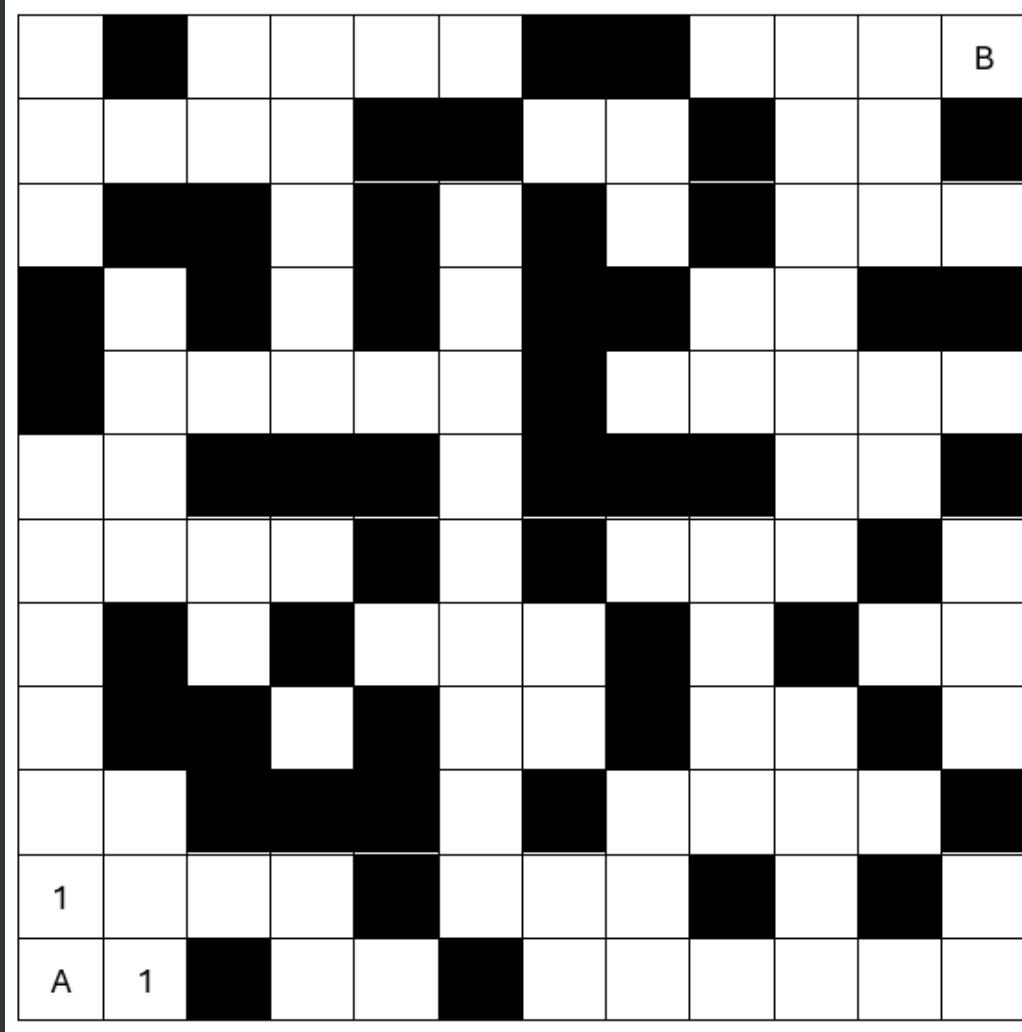
An algorithm is a step-by-step set of instructions or procedures designed to solve a specific problem or complete a specific task. It is a series of well-defined steps that can be followed to achieve a desired outcome. Algorithms are commonly used in computer science and programming to solve complex problems, process data, or perform calculations. They can be written in various programming languages and are fundamental to the functioning of computers and software applications.

25/04/2024, 08:09:11

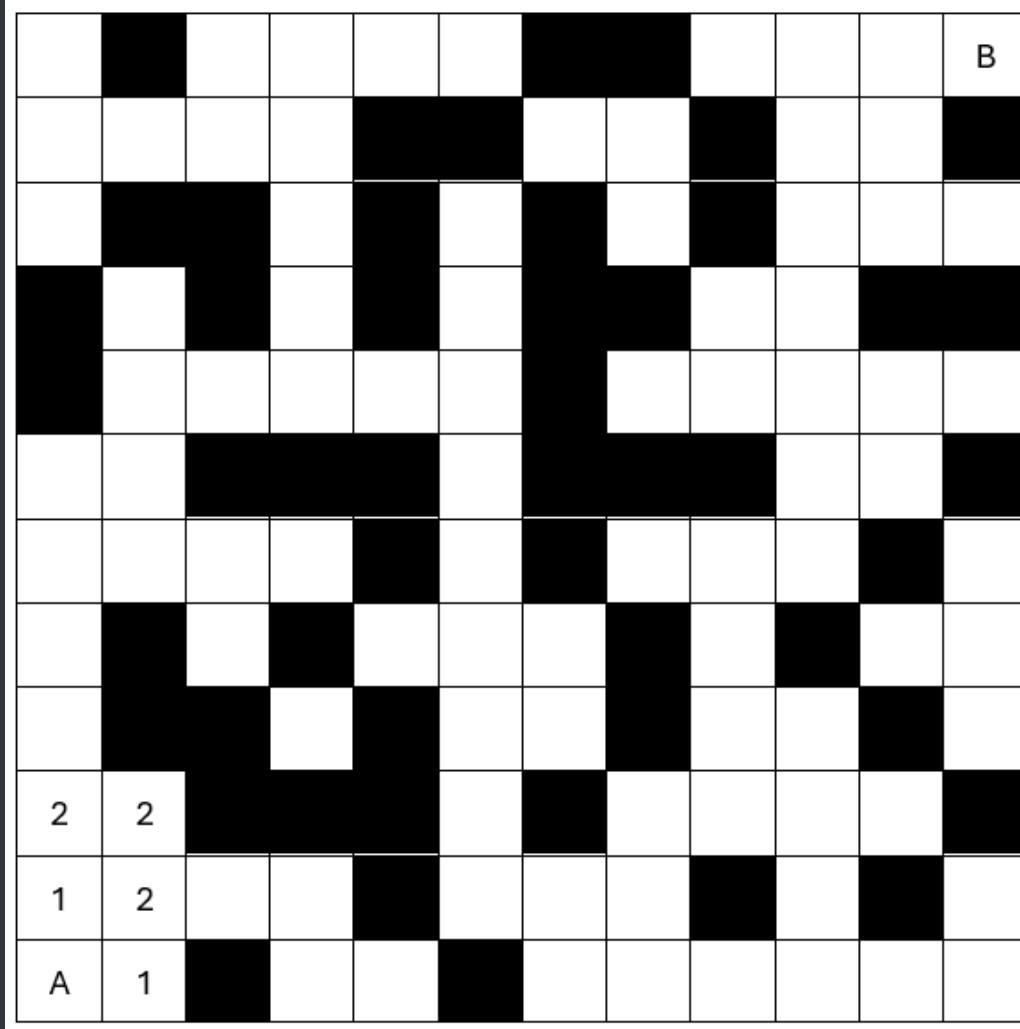
Wait, let me ask you another question.



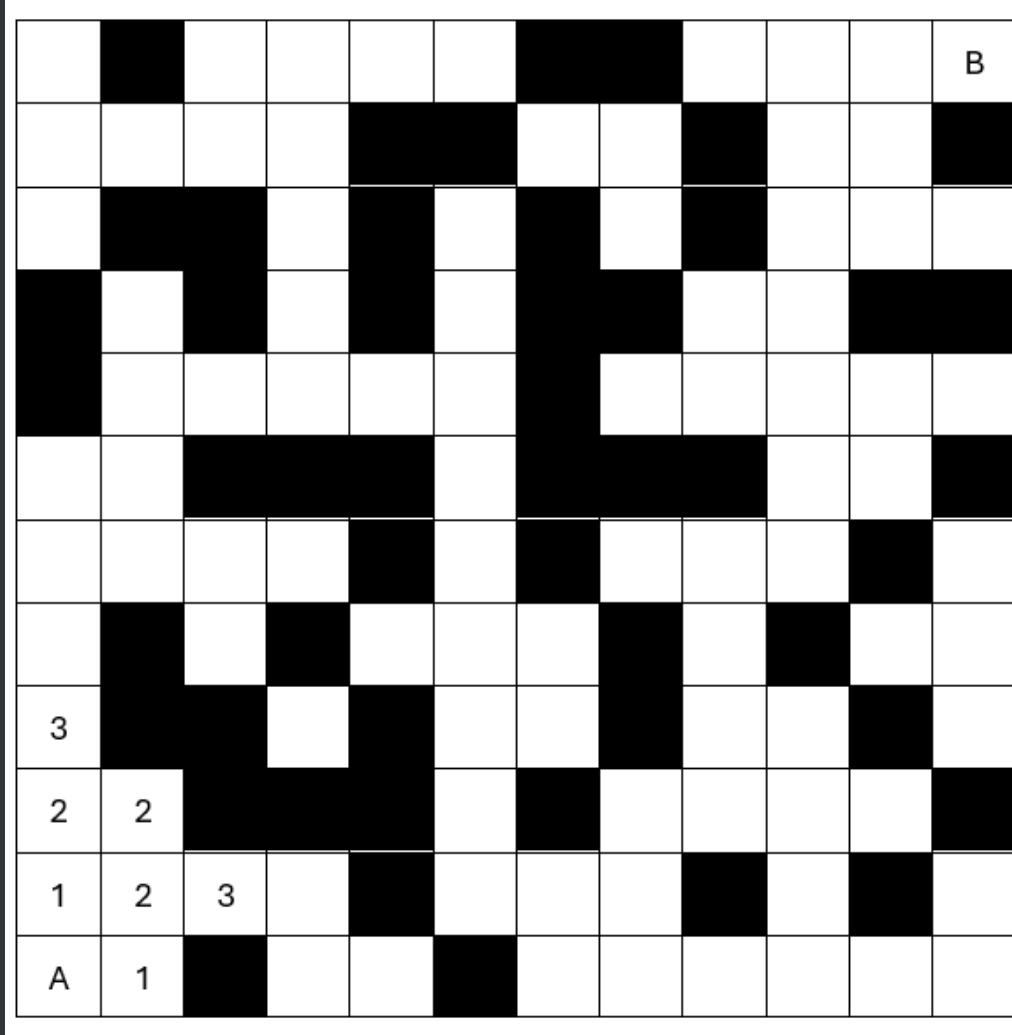
We want to move from A to B. Each step you can move up, down, left or right. But you cannot walk to a black stone. How to find the shortest path from A to B?



These cells can be reached within 1 step.



These cells can be reached within 2 steps.



These cells can be reached within 2 steps.

17		15	14	15	16		33	32	33	34
16	15	14	13					31	32	
17			12		14			30	31	32
	9		11		13		30	29		
	8	9	10	11	12		30	29	28	29
6	7				13			27	28	
5	6	7	8		14		26	25	26	
4		8		16	15	16		24		
3					16	17		23	24	
2	2				17		21	22	23	24
1	2	3	4		18	19	20		24	26
A	1		5	6		20	21	22	23	24
										25

You do this again and again until you reach B.

17		15	14	15	16		33	32	33	34
16	15	14	13					31	32	
17			12		14			30	31	32
	9		11		13		30	29		
	8	9	10	11	12		30	29	28	29
6	7				13			27	28	
5	6	7	8		14		26	25	26	
4		8		16	15	16		24		
3					16	17		23	24	
2	2				17		21	22	23	24
1	2	3	4		18	19	20		24	
A	1		5	6		20	21	22	23	24
										25

You get it!

# Instructor

**Xi Li** (李曦)

Professor of Marketing

Professor of Information and Innovation Management

Director, Asia Case Research Centre

Associate Director, Institute of Digital Economy and Innovation

Ph.D. (Management), University of Toronto

M.Phil. (Operations Research), HKUST

B.E. (Computer Science), Tsinghua University

Email: [xili@hku.hk](mailto:xili@hku.hk)

Office hour: by appointment

Teaching Assistant

Ray Song  
soongray@hku.hk

BTW, Prof. Vincent Zhang and I will be co-teaching an exciting new course on Python programming (offered in Module 5). You can try the course if you want to explore more about programming.

## Three Truths, One Lie

- A) I am married with two kids
- B) I sleep only six hours a day
- C) I suffer from face blindness
- D) I can memorize the map of China



XIAOMI 13 ULTRA | LEICA

2024.09.21 18:10

My Family



With our alumnus

简体中文 ▾



登录 注册

首页 中国 全球 经济 金融市场 商业 创新经济 教育

观点

管理

生活时尚

视频

音频

每日英语

换脑ReWired



Lex专栏 专栏 分析 评论 社评 书评 读者有话说

关注我们 ▾ 工具 ▾



时局  
Present and  
Prospects  
与 前瞻

英国《金融时报》2024年度中国高峰论坛  
2024 FTChinese Annual Forum

链接与增长  
Connectivity and Cultivation  
2024.12.11  
December 11, 2024  
北京  
Beijing

了解详情 →



广告

专栏 明德商论 +关注

## 自愿数据共享，真的可以保护消费者？

李曦：笔者的研究发现，基于自愿的数据分享更可能会推高产品的价格，进一步损害消费者的权益。

李曦

+关注



A bit about my research and its impact

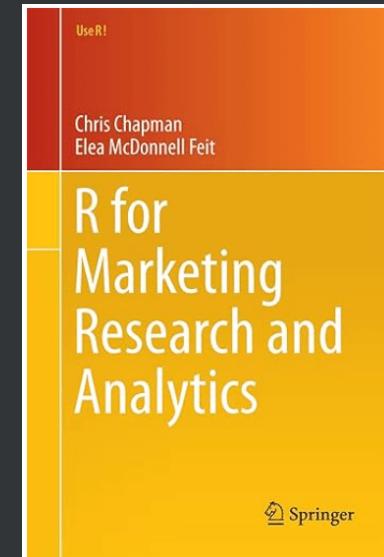
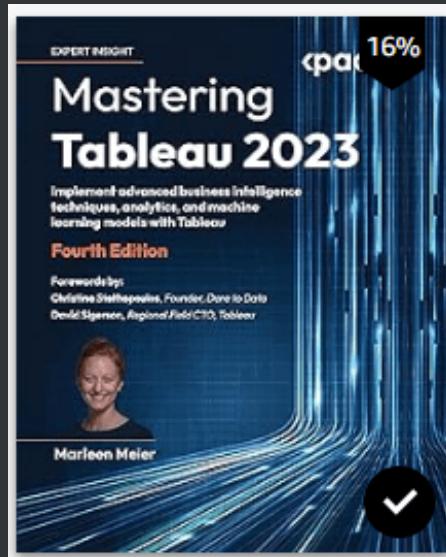
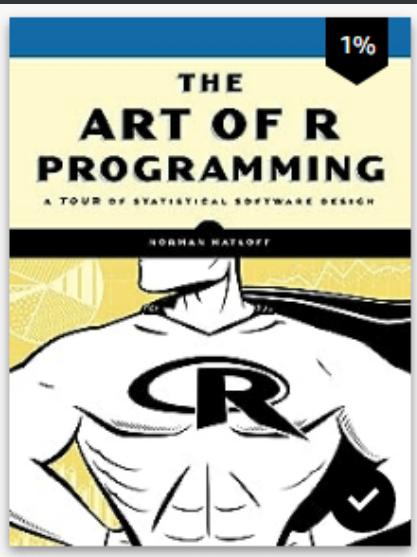
# Teaching Materials

All the course materials will be posted on the  
following website:

[https://ximarketing.github.io/\\_pages/teaching/](https://ximarketing.github.io/_pages/teaching/)

Password is: 2425

No textbooks. However, we suggest some additional reading if you want to explore more.



# Grading

# Grading

The overall grading structure is as follows.

Group work (50%), including:

- Two group data analytics project, 20% each.
- One mini-presentation in class, 10%.

Individual work (50%), including:

- Individual class attendance, 10%.
- Individual class participation, 15%.
- Individual course project, 25%.

# Group Project

Students work in groups on two data projects on the dataset provided by the instructor. You will analyze data using the techniques you learned in (and also outside) the class.

- No presentations for the group project;
- No reports for the group project;
- You only need to submit a few pages slides showing how you analyzed the dataset and the conclusions that you have reached from it.

## Mini Presentation

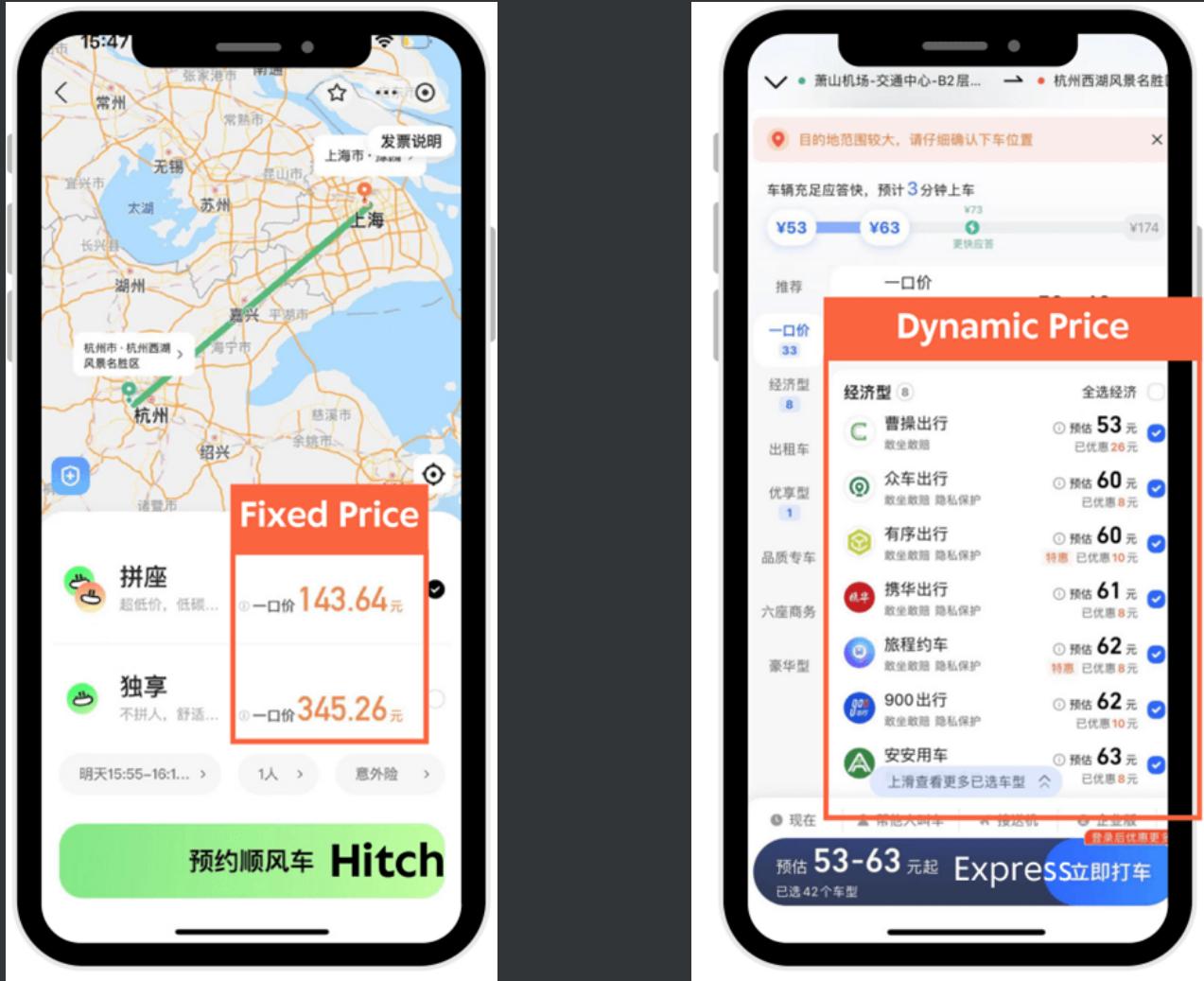
Each group needs to propose an interesting observation from online (or even traditional) marketplaces. You will invite other groups to discuss about the observation and explain the possible causes. The group also offers their own explanation. The presentation only takes around 5 minutes.

I will make a few mini-presentations myself first so that you know how to present.

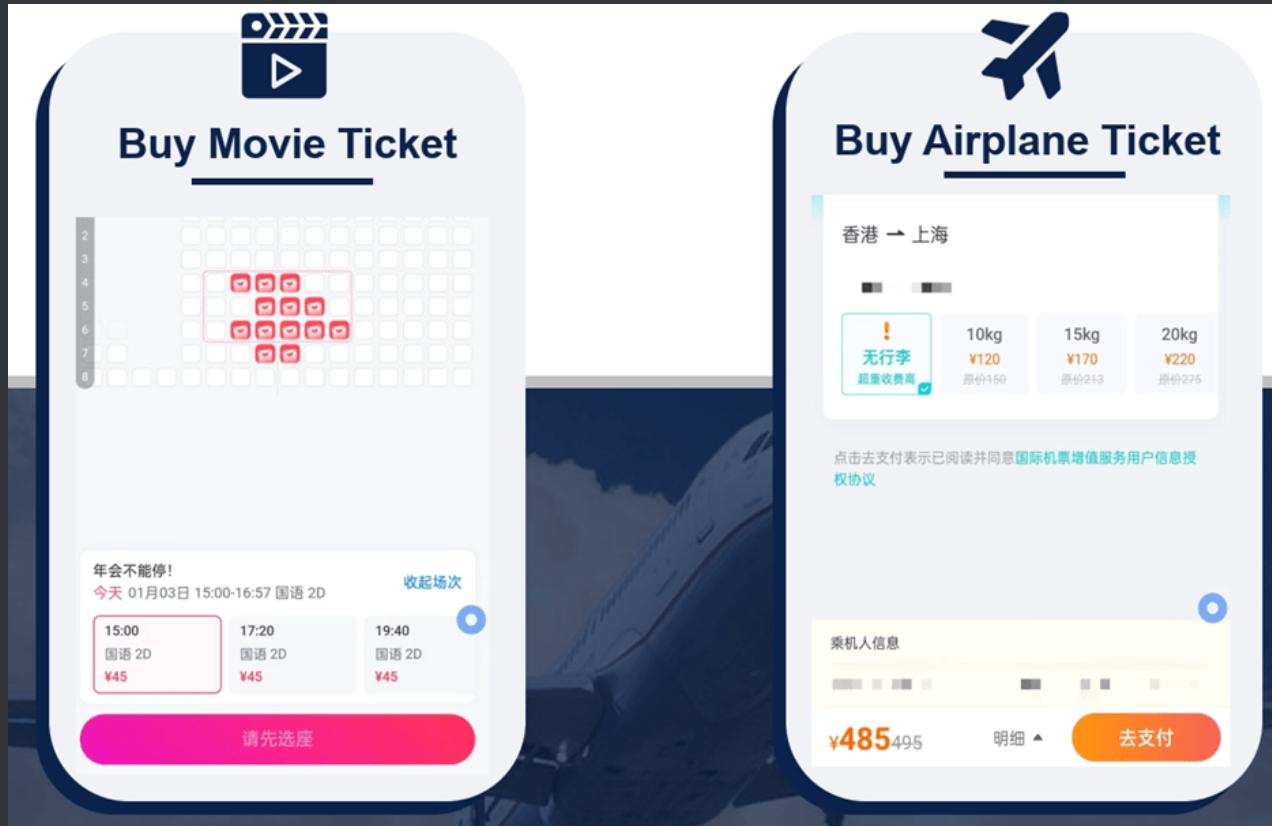
# Mini Presentation: Example



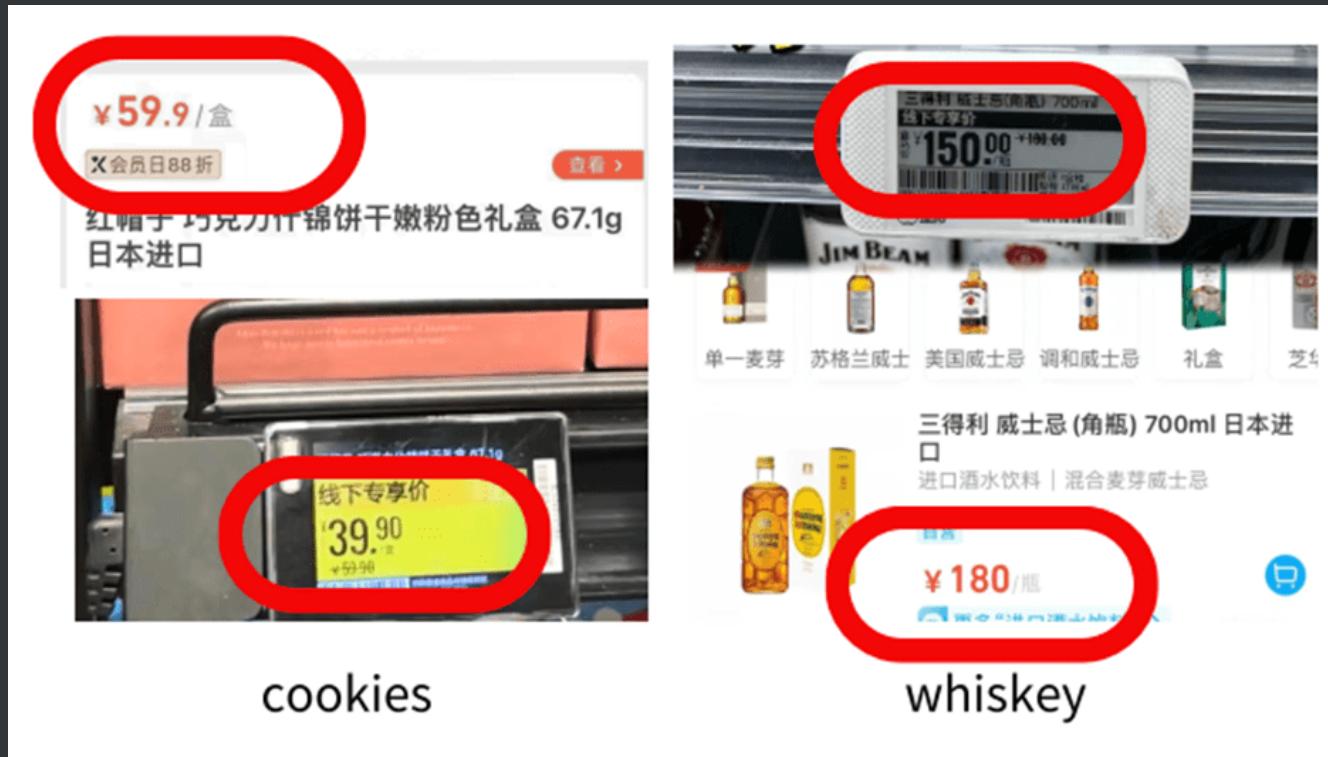
Some online marketplaces display the sales, some display reviews, and some display viewership, why?



Why fixed price (一口价) for Hitch (顺风车) but dynamic price (实时计价) for Express (专车)?



When booking movie tickets, you select seat first and pay later; when booking flights, you pay first and select seats later. But why?



Why does Fresh Hippo (盒马鲜生) charge lower prices offline?

## Individual Attendance

10 points, 1 point for each class.

Scan your HKU ID to check-in each class. If you forget to bring your HKU ID, confirm your attendance manually with the TA during the class.

Otherwise, you miss the point even if you attended the class.

# Individual Participation

Answer questions in class to get your participation points.  
Your points depend on:

- The number of questions answered,
- The hardness of the question, and
- The quality of your answer.

The TA takes notes whenever you answer questions in class.  
If you suspect that TA ignored your participation, check  
with TA immediately, ideally during the next class break.

# Individual Project

Choose **one** of the following two tasks:

- (A) Find a dataset yourself and conduct your own analysis, just like what you do in your group project;
- (B) Find a company that uses data cleverly, and discuss the company's data strategy (no data analysis needed).

Again, no reports, no presentations, only slides.

Anything unclear? Ask me now.

# Overview

# Overview

We have 10 lectures, including

- Today's Introduction
- Two half-class data workshops (totalling 1 lecture)
- Eight regular lectures

# Overview

The topics to be covered include

- Programming with R (fine for beginners)
- Data visualization with Tableau (fine for beginners)
- Web scraping: How to download text and images from the web
- Modeling discrete choice: How do consumers make choice in the supermarket?

# Overview

The topics to be covered include

- Text mining: Getting insights from text
- Causality: Does  $X$  really cause  $Y$ ?
- Tips in data analytics
- Personalized prices and recommendation systems

# Overview

Note that you may have heard about this class from our alumni. However, this year we have an overhaul of the class and the contents are very different from previous years:

- We have brand-new slides this year;
- We include a new data analytics technique, “conditional logit model” in the class;
- We include data analytics techniques for AB tests;
- We updated teaching materials on R and Tableau.

Overall, the class is not harder but becomes more interesting and useful!

# TODO List

## TODO List

We have group assignments, so please form groups with your classmates. **Deadline: Dec 3, 23:59.**

- Each group can have up to 8 students.
- Create a name for your group.
- Submit your group information to your TA.
- Let me (or the TA) know if you do not have a group.

If you do not have a group and do not let us know, we assume that you prefer to work as a one-person group.

## TODO List

Here are some past group names for your reference (listed in alphabetical order):

- 25 hours
- CEOs
- Data Geeks
- Data Killers
- Doctor Malou
- GGBond
- Gold Miners
- Sense 8

## TODO List

Our class uses R and Tableau for teaching. Please install R and RStudio on your laptop and bring it with you for the next class. You can

- Download R [here](#)
- Download RStudio [here](#)

Note that your installation path should not contain any non-English letters, otherwise you will be unable to use some functions.

安装路径必须为纯英文，否则某些功能无法正常使用！