

Algorithms, Big Data, and Online Marketplaces

~~Algorithms, Big Data, and Online Marketplaces~~

Algorithms, AI, and Market Analytics

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

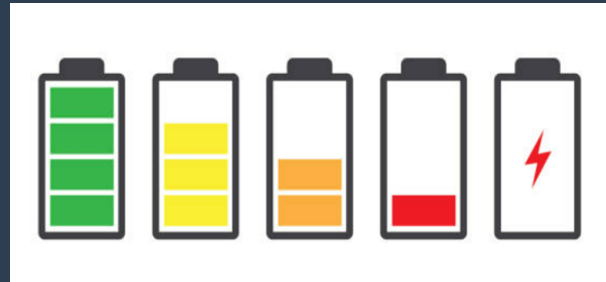
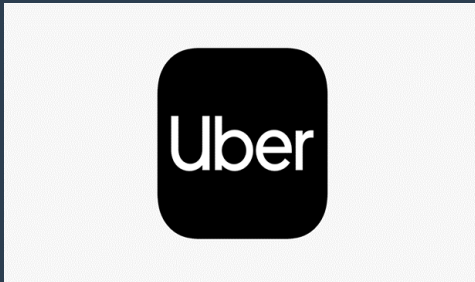


What are the wealthiest firms in the world?

What were the wealthiest firms in the world in 2015?

<https://www.youtube.com/embed/BQcDg5iSwjM?enablejsapi=1>

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.



How many of you have Taobao / JD / PDD installed on your phone?

Do you know what...

A shopping app will check if you have other shopping apps installed. If you have PDD on your phone, you will probably get better deals on Taobao!

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

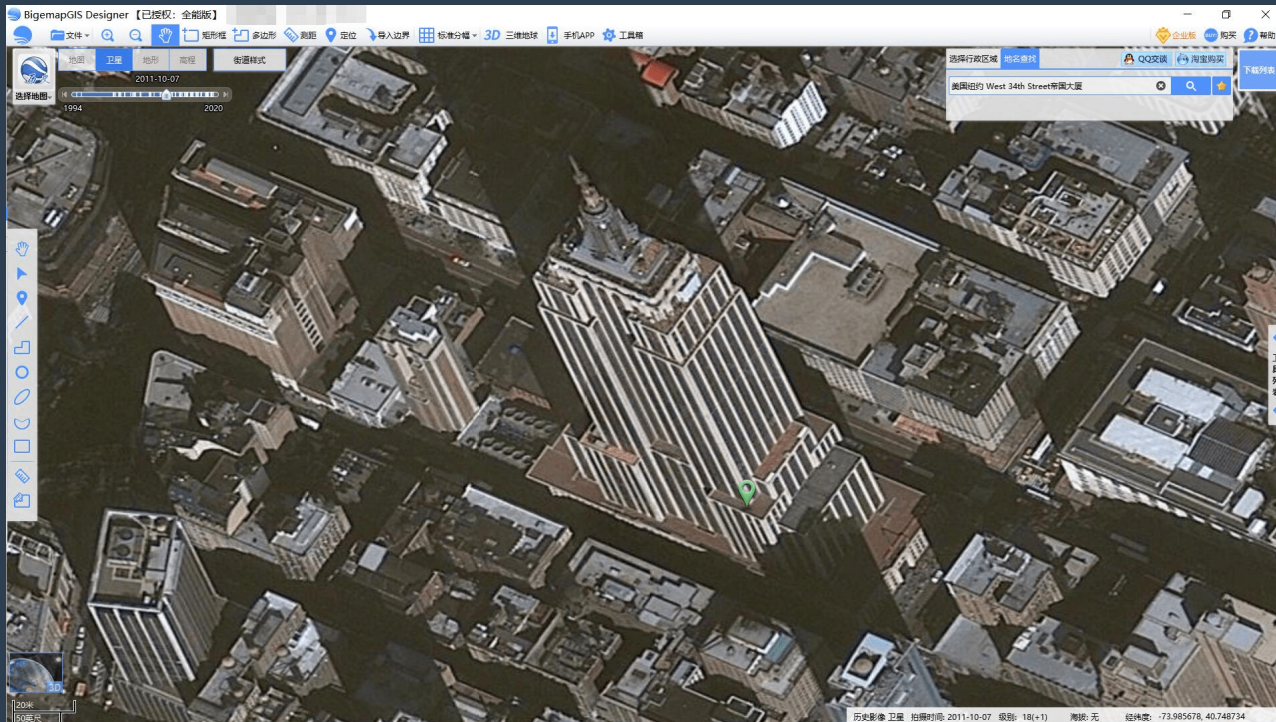
Pinar Yildirim, Christophe Van den Bulte

Marketing Department, The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104
{pyild@wharton.upenn.edu, vdbulte@wharton.upenn.edu}

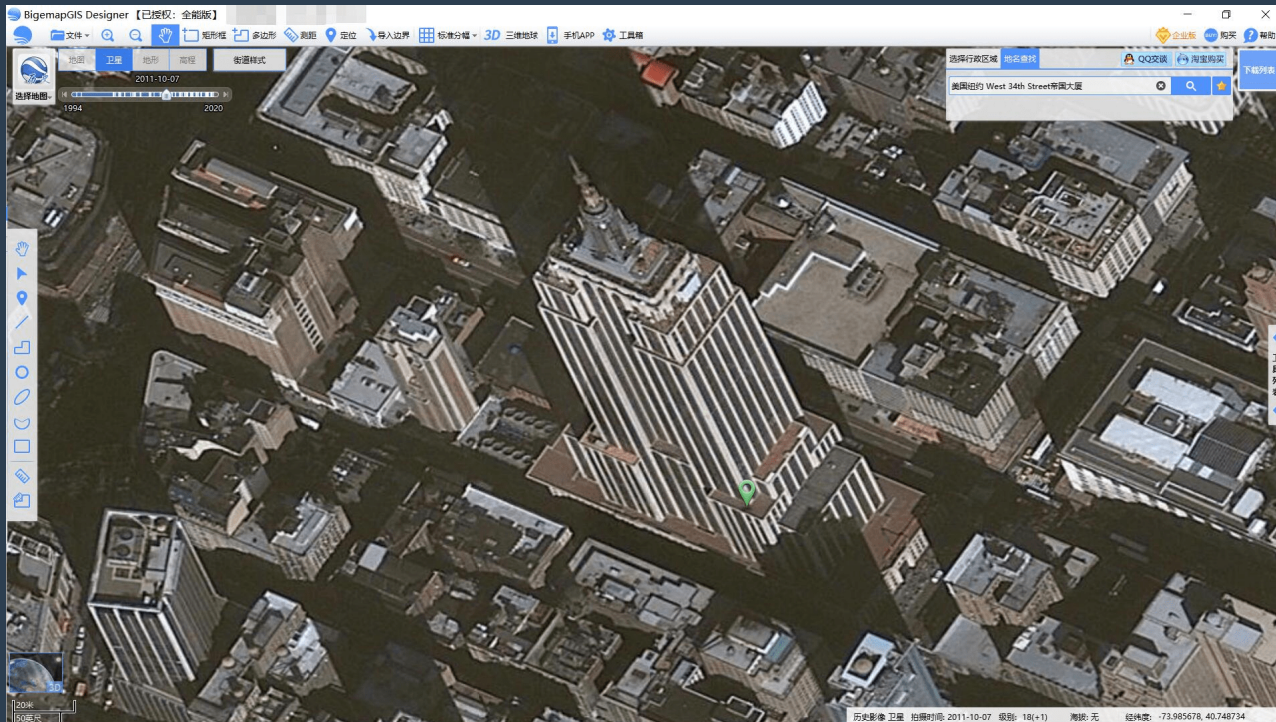
Chrysanthos Dellarocas

Information Systems Department, Questrom School of Business, Boston University, Boston, Massachusetts 02215, dell@bu.edu

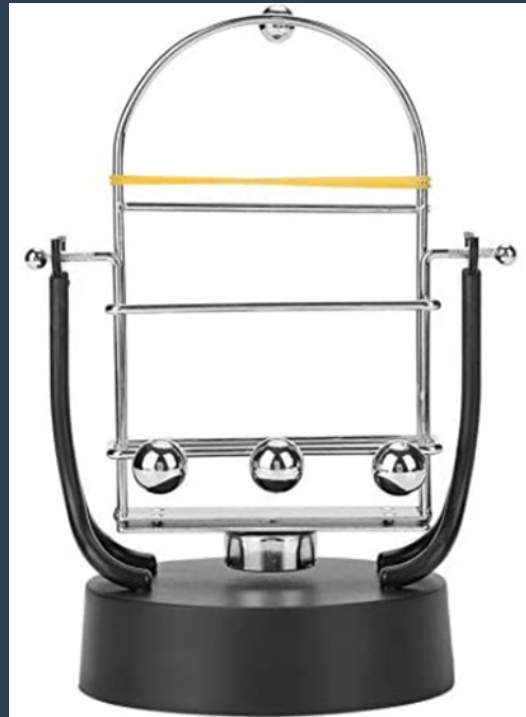
US retailers Target and Kroger both purchase real-time satellite images from RS Metrics. But why?



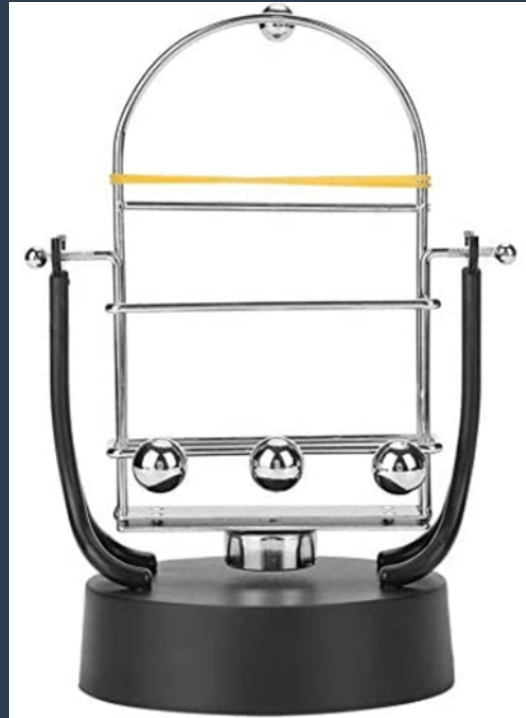
They monitor the competitors' carpark to estimate the competitors' performance and revenue.



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

Replying to @mbrennanchina

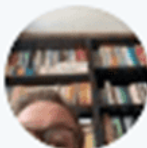
Aka adversarial poisoning attack on data-driven algorithms...



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



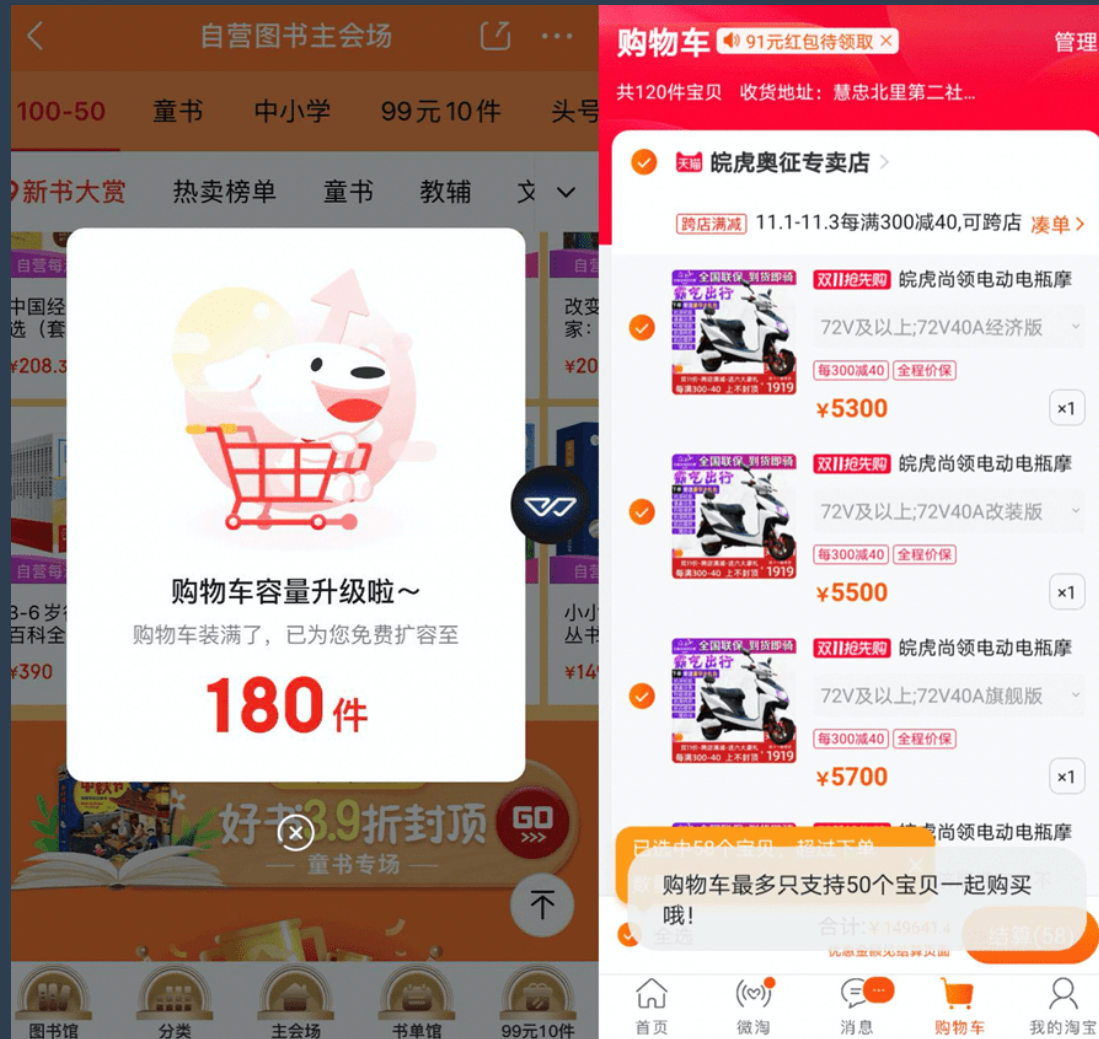


Netflix: How did it know I was bi before I did?

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.

What is the biggest difference between an offline matchmaker (红娘) and a dating platform (婚恋网站)?

Both profit from matchmaking, but their business models differ:

- A matchmaker charges you a fixed amount when you successfully get married;
- A dating platform makes money from your monthly subscription fees. It does not know if you eventually get married with your online partner or not.

So what?

- A matchmaker wants you to get married as soon as possible! In this case she will make immediate profits.
- A dating platform **loses future revenue** from you when you get married and exit the market! It hopes that you can stay at the platform as long as possible!

A dating platform does not need a perfect matching algorithm, it wants its matching to be OK but not perfect.

What's an algorithm?

We signed up for a course on algorithms, but let's be honest
—what is an algorithm supposed to be?

Don't ask me. Ask AI!

What is an algorithm?

An **algorithm** is a **step-by-step procedure** or a **precise set of rules** for solving a problem or completing a task.

More concretely:

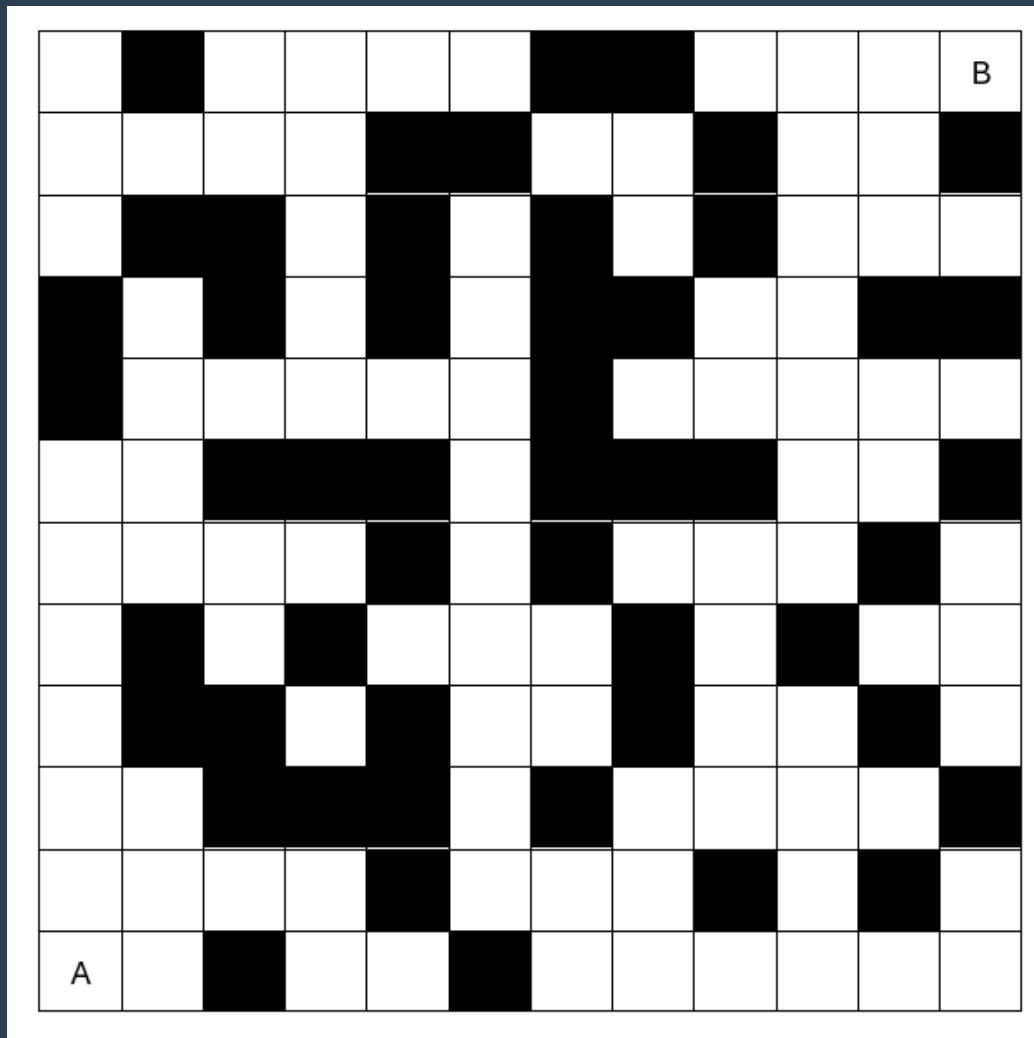
- It takes **input** (data you start with).
- It follows a **defined sequence of steps**.
- It generates **output** (a result).

Algorithms can be expressed in many forms — natural language, flowcharts, pseudocode, or actual computer programs — but the key is that the steps must be clear enough that someone (or a computer) could follow them unambiguously.

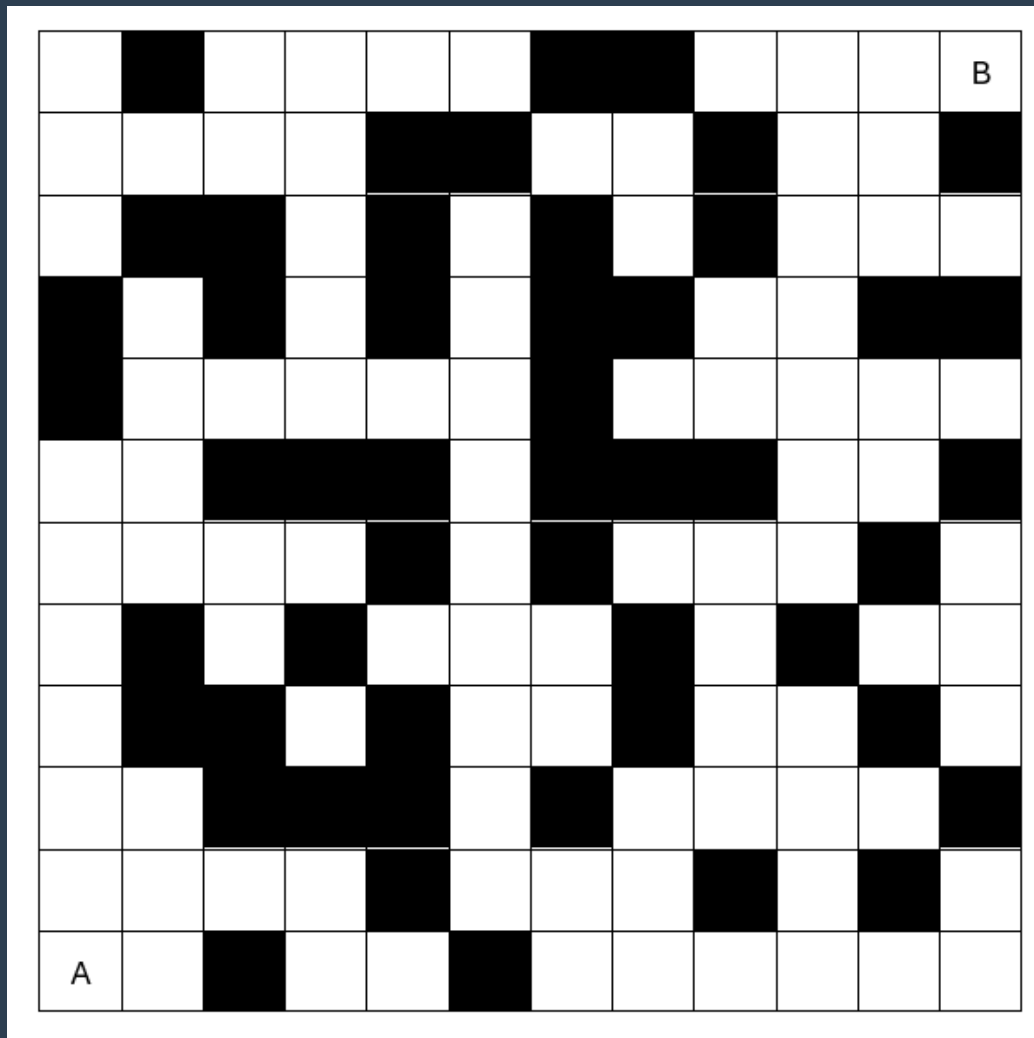
Example:

A recipe is an everyday example of an algorithm: it lists ingredients (input) and instructions (steps) to produce a dish (output).

Wait, let me ask you another question.

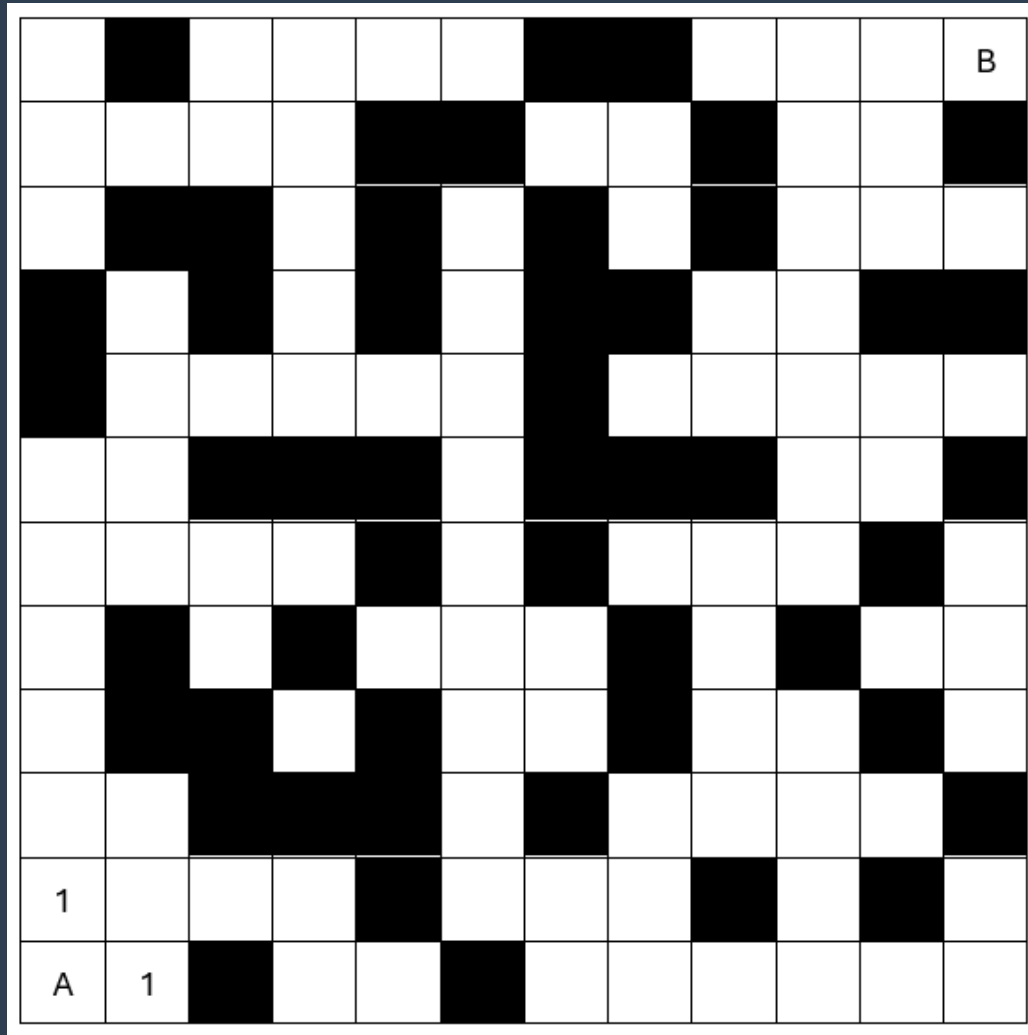


We want to move from A to B. Each step you can move up, down, left or right. But you cannot walk to a black block. How to find the shortest path from A to B?

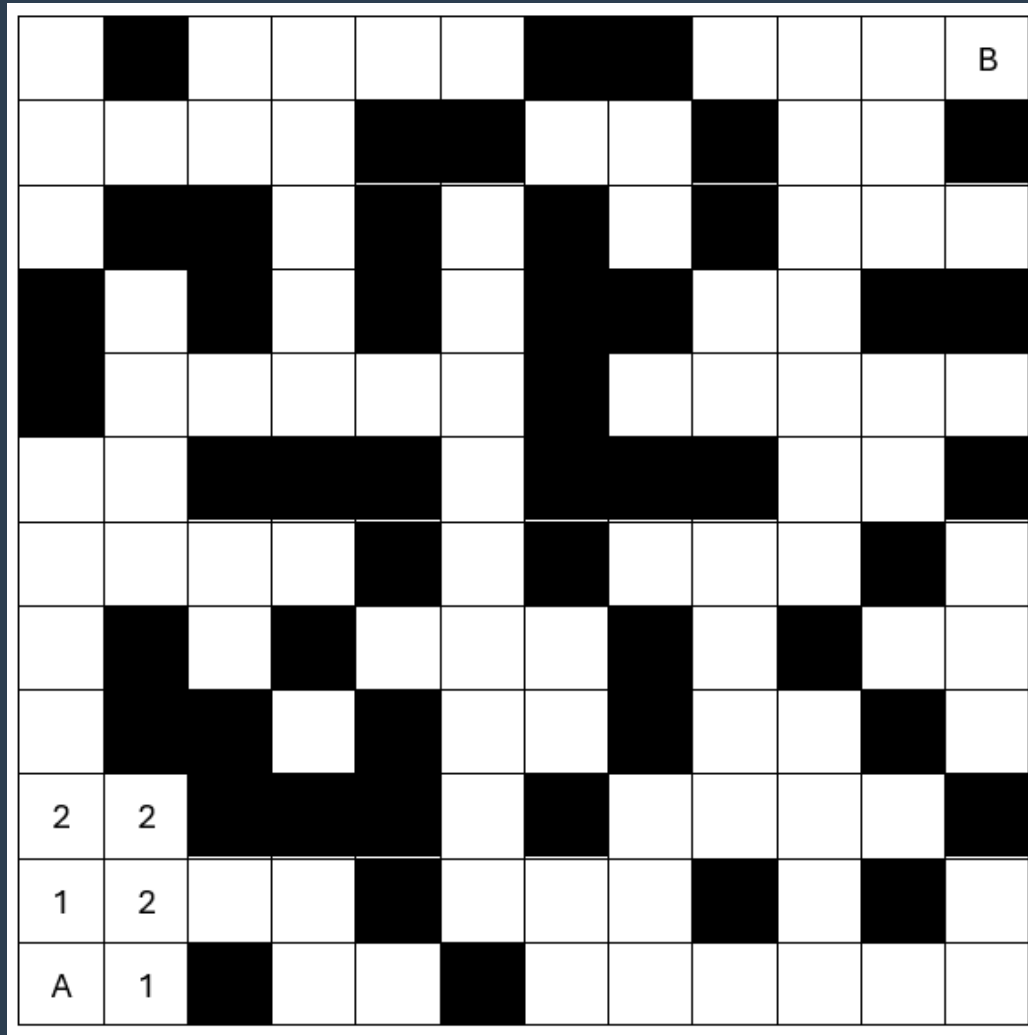


Input: A map

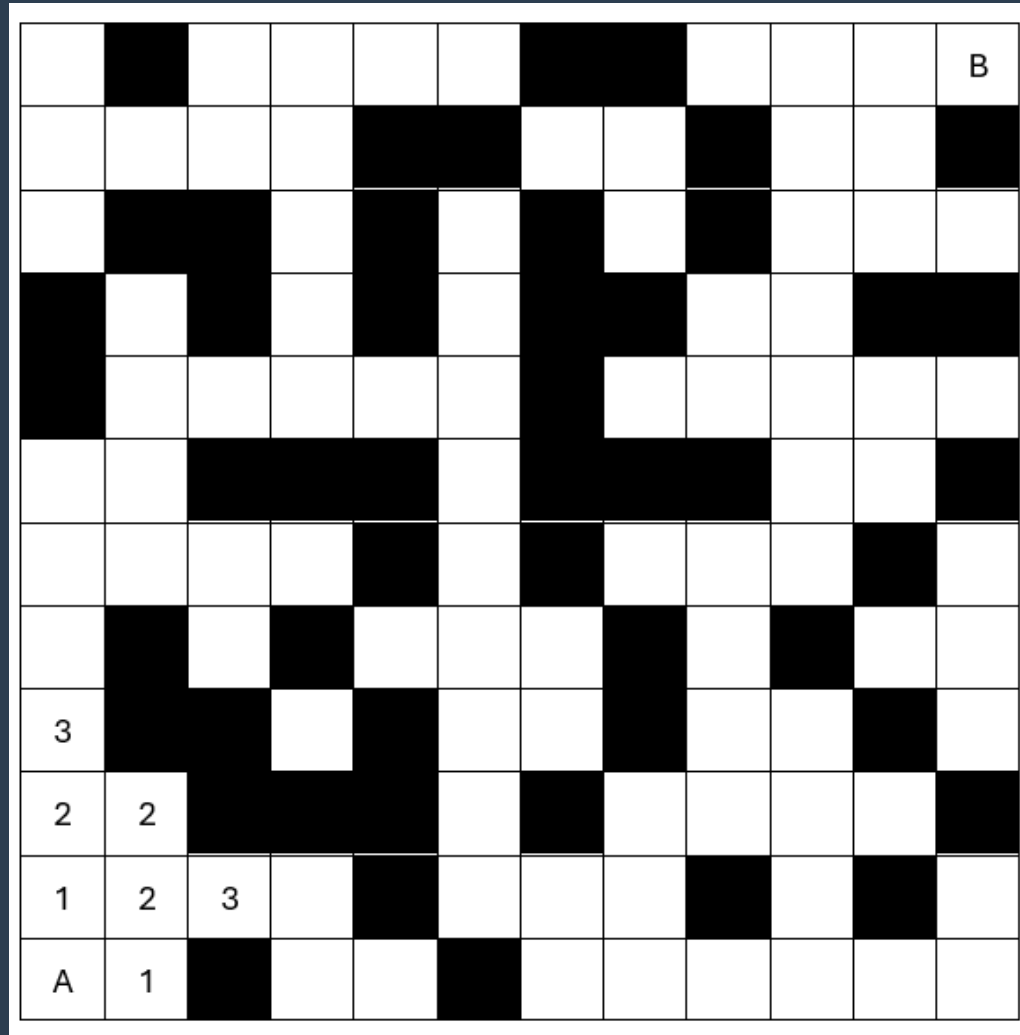
Output: The shortest path from A to B



These cells can be reached within 1 hop.



These cells can be reached within 2 hops.



These cells can be reached within 3 hops.

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	30
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	30
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 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

Office hour: by appointment

Teaching Assistant

Kevin Zhang
kevinzyx@hku.hk

Kevin is also an alumnus of our MSMK programme.

BTW, Prof. Vincent Zhang and I will be co-teaching a 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





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](#)



电商下半场 如何从“流量至上”走向“品质为王”？

2025年11月19日19:46 | 来源：人民网

T_T 小字号



对话现场。人民网记者 任峰涛摄

Teaching Materials

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

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

Password is: 26ATC

Grading

Grading

The overall grading structure is as follows.

Group work (50%), including:

- One group data analytics project, 20%.
- **One group AI project, 20% (new!) -- it replaces a data project.**
- One mini-presentation in class, 10%.

Individual work (50%), including:

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

Group Data Project

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

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

Group AI Project (New!)

Students work in groups on one AI project. You will use a large language model to generate some interesting and innovative content.

- **No presentations** for the group project;
- **No reports** for the group project;
- You only need to submit the prompts you used to communicate with AI and the output of your AI.

Mini Presentation

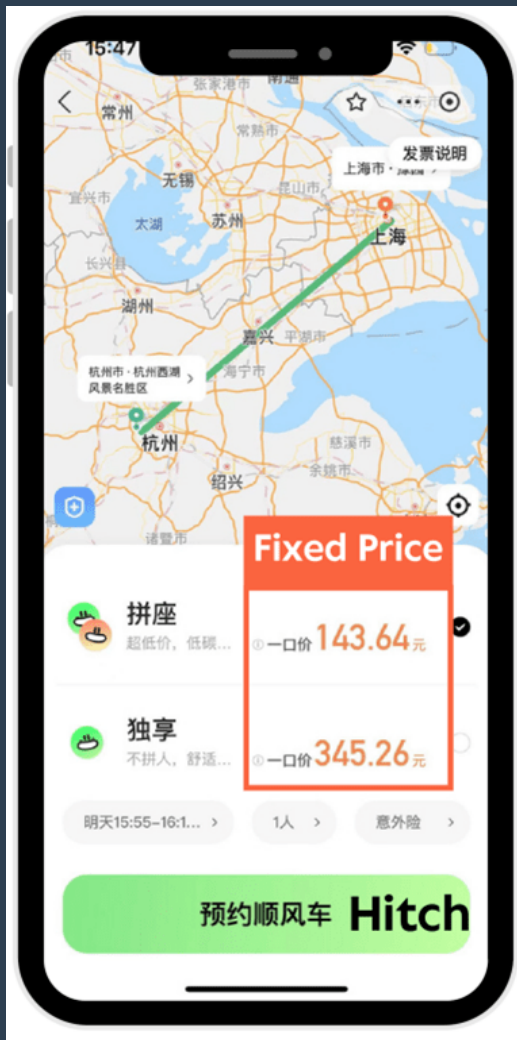
Each group needs to propose an interesting observation from online (or even traditional) marketplaces. You will invite other groups to discuss about the observation and explain the possible causes. The group also offers your own explanation. The presentation only takes around 5 minutes.

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

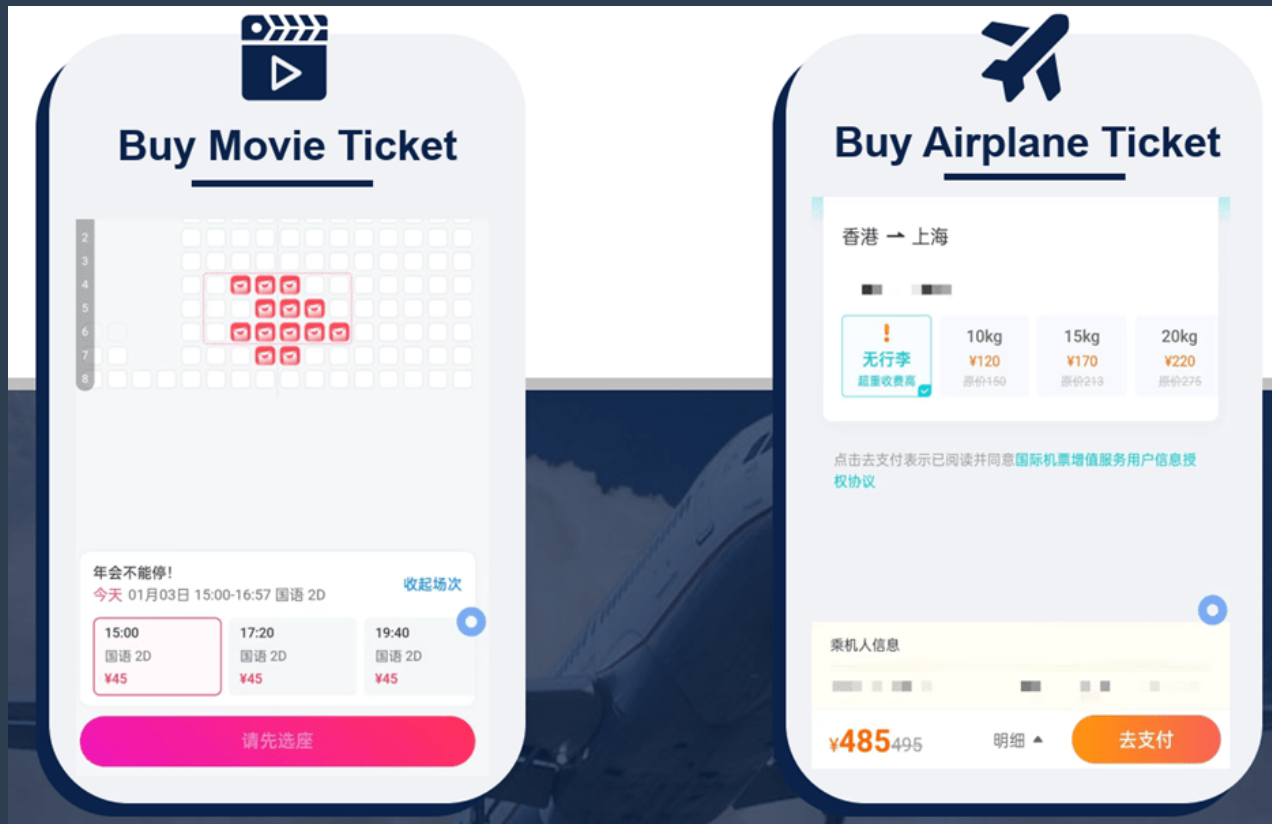
Mini Presentation: Example



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



Why fixed price (一口价) for Hich (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?

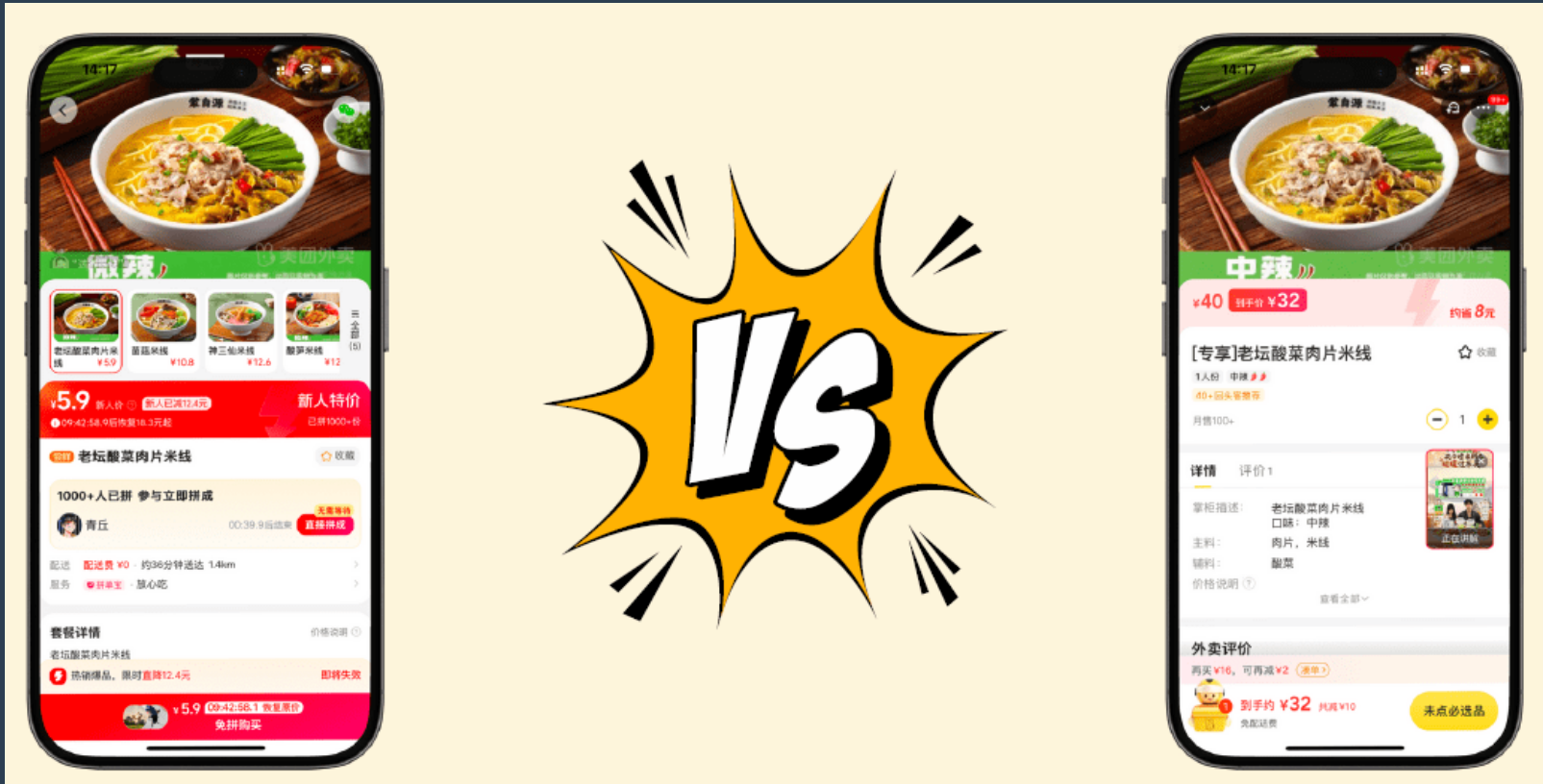


cookies

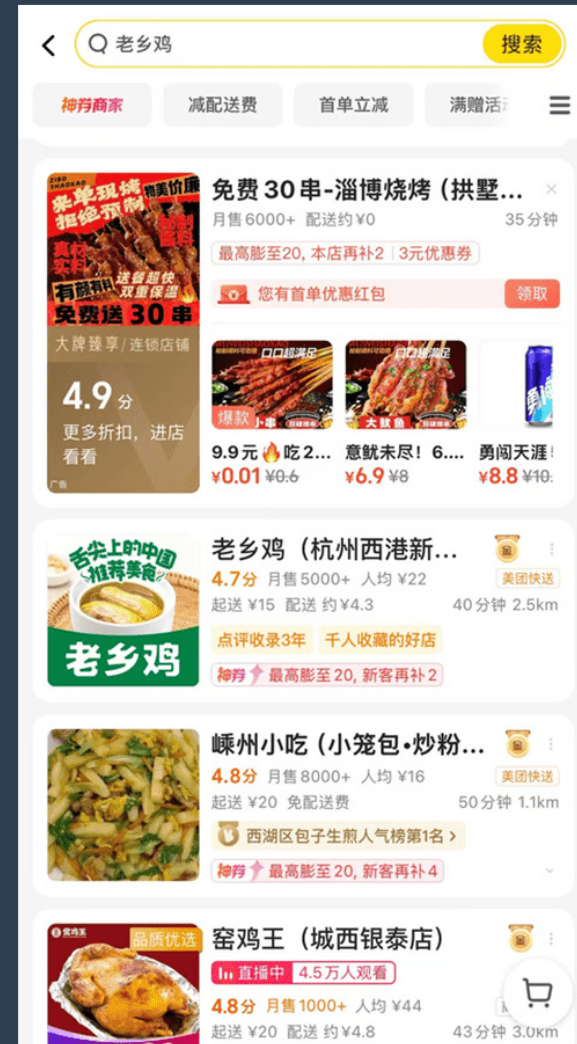


whiskey

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



Why is the same product cheaper on Pinhaofan (拼好饭)?

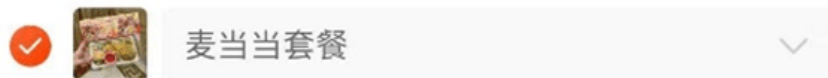
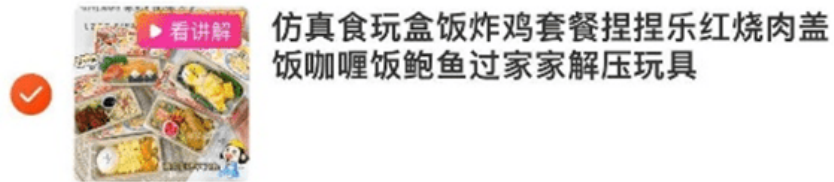


Meituan ranks restaurants while Pinhaofan ranks products.
Why?

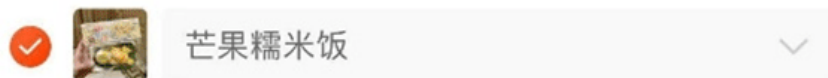


Why does the same movie receive so different ratings on Douban and Maoyan?

✓  义乌市七界贸易有限公司 >



¥8.00 — 1 +



¥8.00 — 1 +



有2种商品因不符合商家混批下单要求, 未被勾选

⊖ 全选

优惠: ¥5.00 共计: ¥14.80

2种2件 | 含运费 金额明细 ^

结算

新

首页

年

会员店

齐⁵

消息

奔²

采购车

富

我的

On 1688, shipping cost is listed on product page.



On Taobao, shipping cost is listed on the checkout page only.

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. This is out of my control (I don't care about your attendance myself).

Individual Participation

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

- The number of questions answered,
- The difficulty of that question, and
- The quality of your answer.

The TA takes notes whenever you answer questions in class. If you believe your participation was overlooked, please check with the TA right away—ideally during the next class break.

Individual Project

Choose **one** of the following three 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).

(C) Use AI to generate some content (**tentative**, TBC).

No presentations. Submit slides for (A), report for (B) or your output and prompts for (C).

Anything unclear? Ask me now.

Overview

Overview

We have 10 lectures, including

- Today's Introduction
- One half-class data workshop and one guest lecture (to be confirmed)
- Eight regular lectures

Overview

The topics to be covered include

- Data visualization with Tableau (fine for beginners)
- Programming with R (**with the help of AI**; 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

- Large language models: prompt engineering and RAG
- AI Agents
- Causality: Does X really cause Y ?
- 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 new AI contents this year.
- We add AI components to classical content.
- The slides are all updated.

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

TODO List

TODO List #1

We have group assignments, so please form groups with your classmates. **Deadline: Dec 13, 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 #1

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 #2

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

- Download R [here](#)
- Download RStudio [here](#)
- Download Tableau Public [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！

TODO List #3

Join Github Copilot ([Free Education Version](#)). You need to sign up for a Github copilot account.

Follow the instructions [here](#) to add Copilot to your RStudio. Thanks to our TA for writing the detailed instructions.

TODO List #4

You probably need a VPN to visit the official **ChatGPT website** [the HKU version is problematic]. You can choose any VPN (many provide a free trial period).

If you cannot visit the official ChatGPT website, you may use other tools such as Deepseek and Gemini instead, but the quality of the answers varies a lot.