Algorithms, Big Data and Online Marketplaces

Welcome to the course!

Scan the QR code to join a survey

Code:

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?

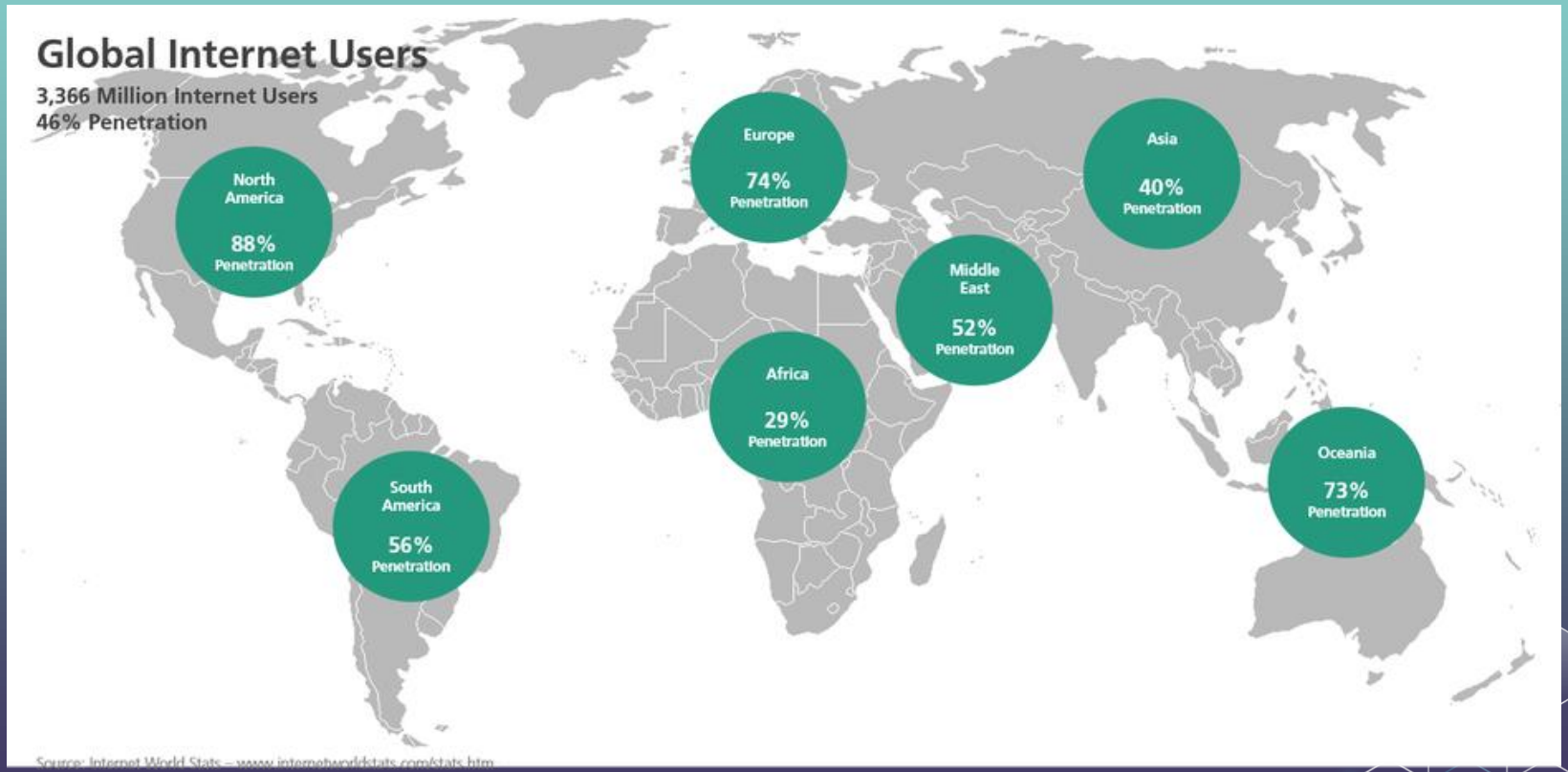
What is the Internet penetration rate in HK/Mainland China?

There were 6.92 million internet users in Hong Kong in January 2021. There are still 8% that do not use the Internet.

In mainland China, about 70% of the population used the internet.

The Internet penetration rate of the whole world is 59.5%.

Global Internet Penetration



Reasons for not using the Internet

No interest or did not think the internet was relevant (34%)

Too difficult to use or “too old to learn” (32%)

The expense of internet service or owning a computer (19%)

Do You Know the Answer?

Among every $100 Hongkongers spend on retailing, how many dollars are spent online?

The 2019 FIS Retail Global Payments Report shows Hongkongers spend in online retail shopping at 4% of overall purchases.

This is lowest worldwide (mainland: 24%; world average: 9.7%).

Do You Know the Answer?

In China, a consumer spends, on average, _____ RMB on Taobao and Tmall (“淘系”) every year.

According to Alibaba’s financial report in financial year 2020, an average consumer spends 9076 RMB on its online platforms.

On JD (京东) and PDD (拼多多), the number is 5761 and 1720 RMB, respectively.

Do You Know the Answer?

On average, how much time does a Hong Kong/Mainland teenager spend on their smartphones?

The consulting firm TNS said the average millennial aged 16 to 30 in the city spends 2.8 hours a day on their mobile devices.

In mainland China, that is 3.9 hours. In Japan, the number is 1.6 hours.

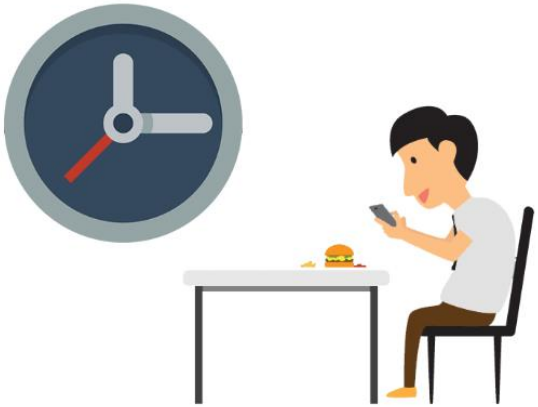
Avg. Use Per Day

2:51 Hrs (171 Mins) per day

The average time spend on smartphones is 2hrs 51mins a day

4:33 Hrs (171 Mins) per day

The average time spend on smartphones AND tablets is 4hrs 33mins a day



18-29 Year Old Interactions

22% check their phone every few minutes out of smartphone owners surveyed

51% check their phone a few times per hour out of smartphone owners surveyed

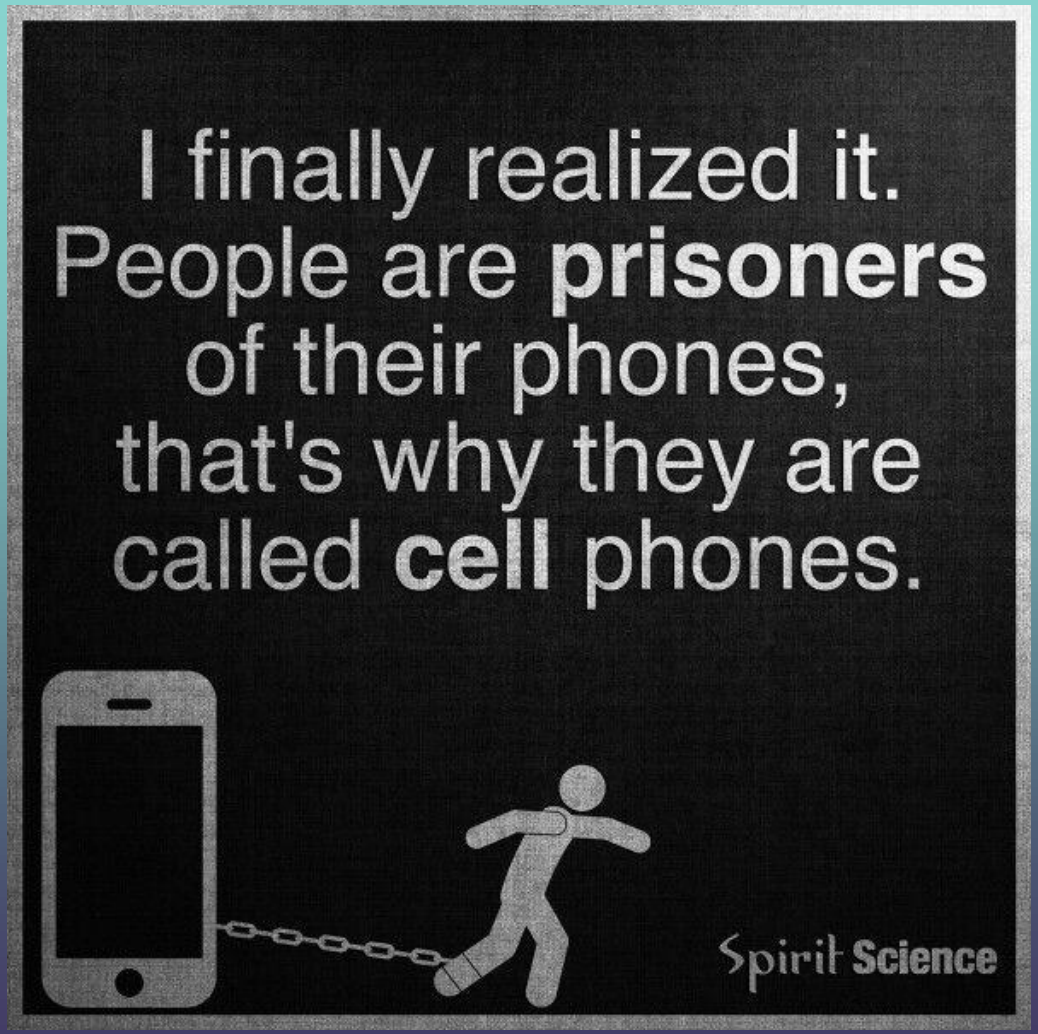
Social Interactions

1:16 Hrs (76 Mins) per day

The average user spends 1hr 16mins a day on the top 5 social media apps

2,617 Daily taps, clicks swipes

The average user will tap, swipe, click their phone 2,617 times a day

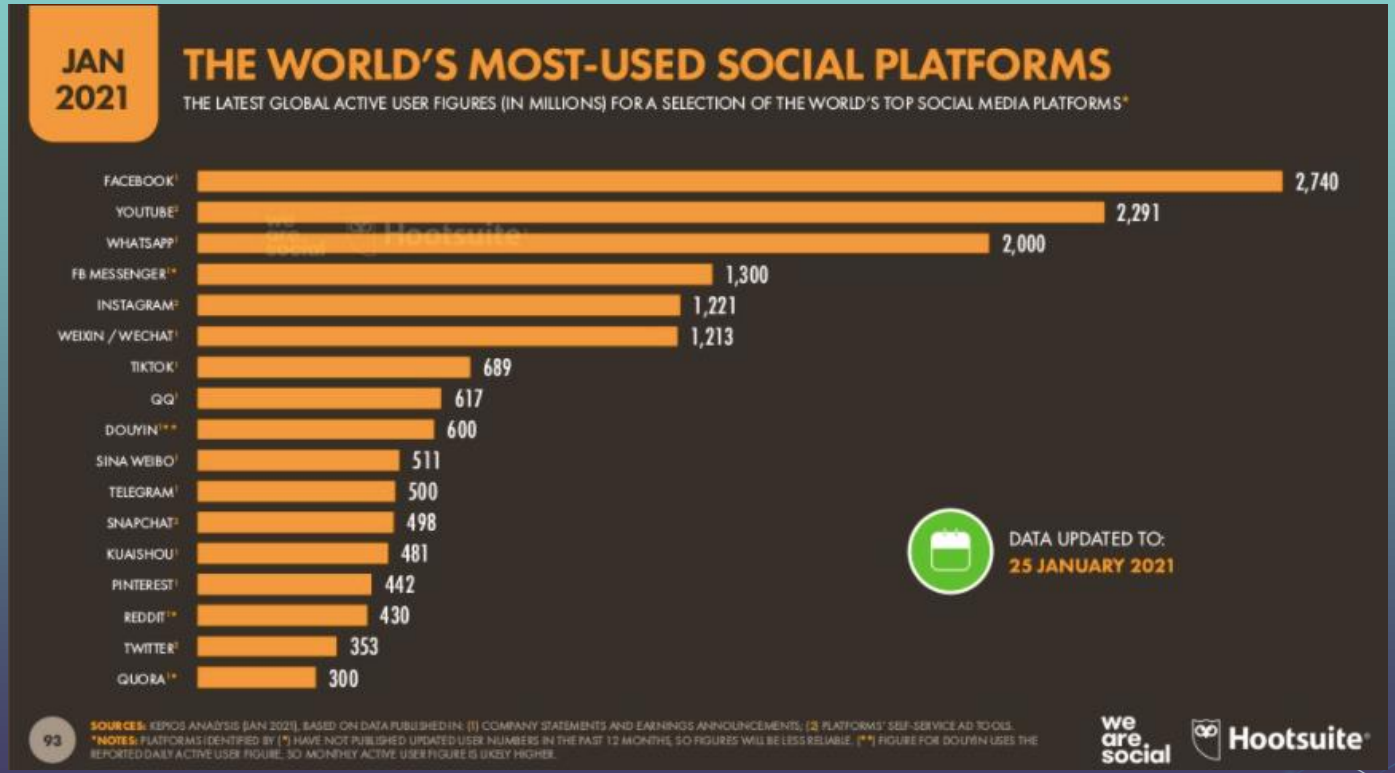

I finally realized it.
People are **prisoners**
of their phones,
that's why they are
called **cell** phones.
Spirit Science

Do You Know the Answer?

What are the most popular social media platforms in Hong Kong?

As of March 2021, Facebook is the most popular social media platform in Hong Kong, followed by Twitter, Pinterest, YouTube and Instagram.

This is slightly different from the world data



Do You Know the Answer?

How many couples meet their partner online?

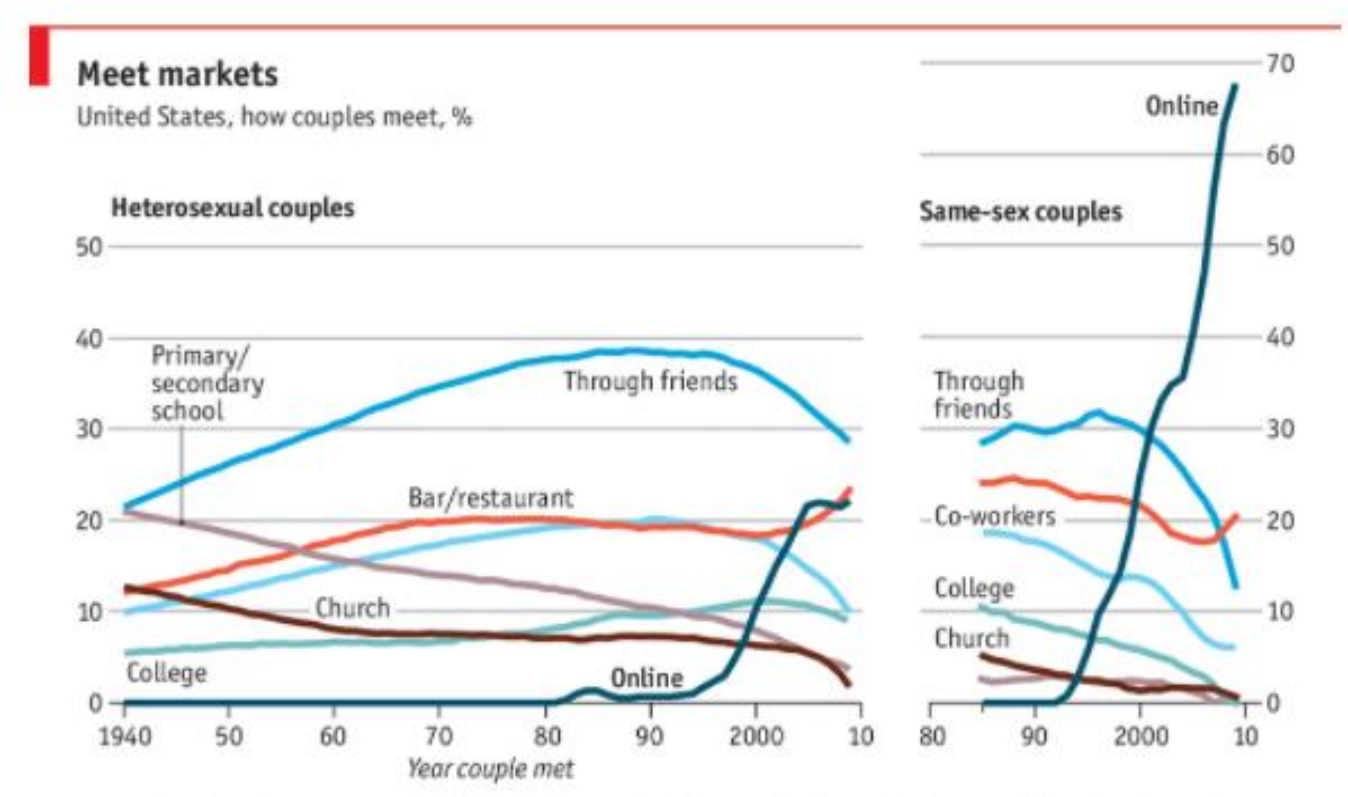
According to a Stanford University research project, by 2017, 39 percent of heterosexual couples reported meeting their partner online.

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)



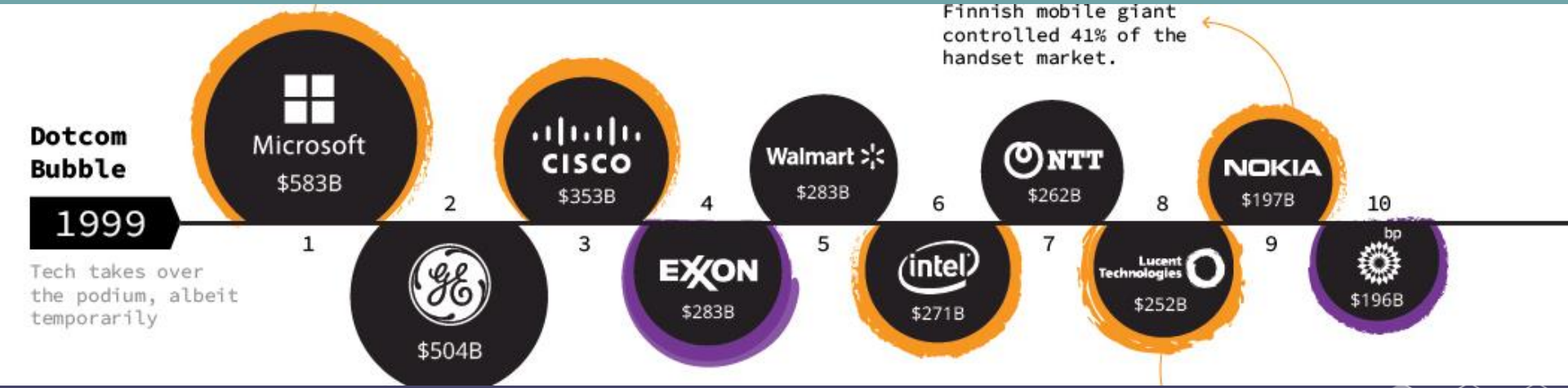
The Internet is changing our lives

Internet creates wealth

Question: which companies have the highest market capitalization (市值) worldwide?

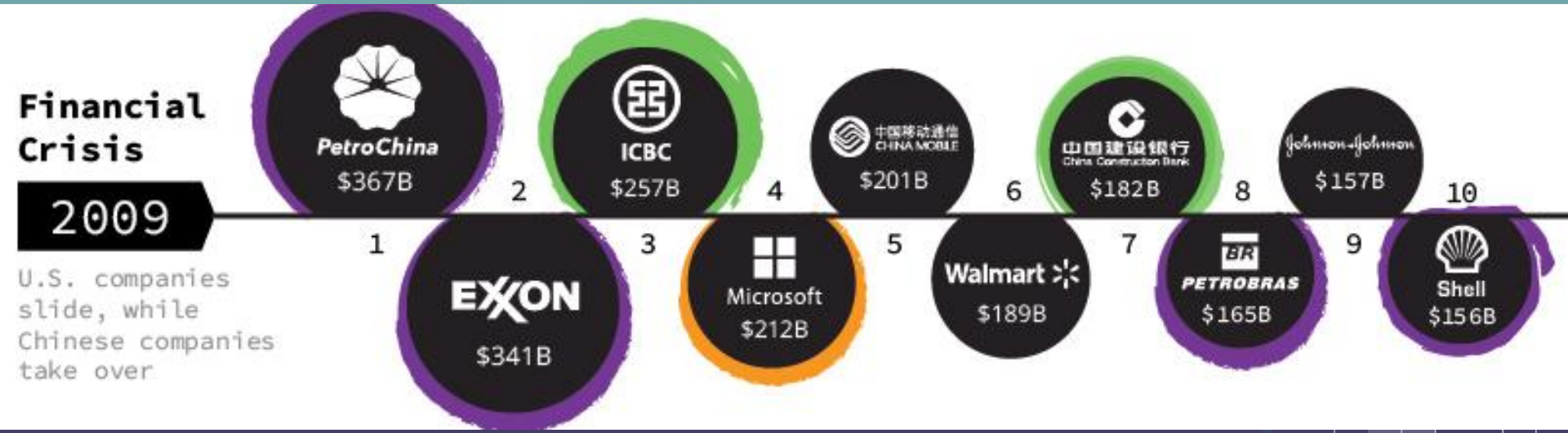
Internet creates wealth

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 2021, they are

^1	1	Apple AAPL	$2.634 T	$160.55	1.70%		USA
v1	2	Microsoft MSFT	$2.576 T	$343.11	0.54%		USA
	3	Alphabet (Google) GOOG	$1.984 T	$2,999	-0.50%		USA
	4	Saudi Aramco 2222.SR	$1.960 T	$9.81	-0.67%		S. Arabia
	5	Amazon AMZN	$1.864 T	$3,677	-0.53%		USA
	6	Tesla TSLA	$1.141 T	$1,137	3.71%		USA
	7	Meta (Facebook) FB	$960.54 B	$345.30	1.95%		USA
	8	NVIDIA NVDA	$821.98 B	$329.85	4.14%		USA
	9	TSMC TSM	$644.42 B	$124.26	0.71%		Taiwan
	10	Berkshire Hathaway BRK-A	$628.07 B	$423,374	0.16%		USA
	11	Tencent TCEHY	$606.41 B	$63.20	1.28%		China
	12	JPMorgan Chase JPM	$475.56 B	$160.92	-1.31%		USA


Biggest
Companies
Mar 2011

Most Attractive Employers in China

10. Meituan
9. Huawei Technologies
8. Amazon
7. Didi
6. Tesla
5. Nio (蔚来)
4. Fosun (复星)
3. ByteDance
2. Baidu
1. Alibaba

These are the best companies to work for in China in 2019, according to LinkedIn

Most Attractive Employers in HK

10. J.P. Morgan
9. UBS Bank
8. Morgan Stanley
7. Richemont 历峯集团
6. HK Electric
5. LVMH
4. Swire Properties 太古地產
3. The Hong Kong & China Gas Company (Towngas)
2. Cushman & Wakefield 戴德梁行
1. Hong Kong Jockey Club

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:

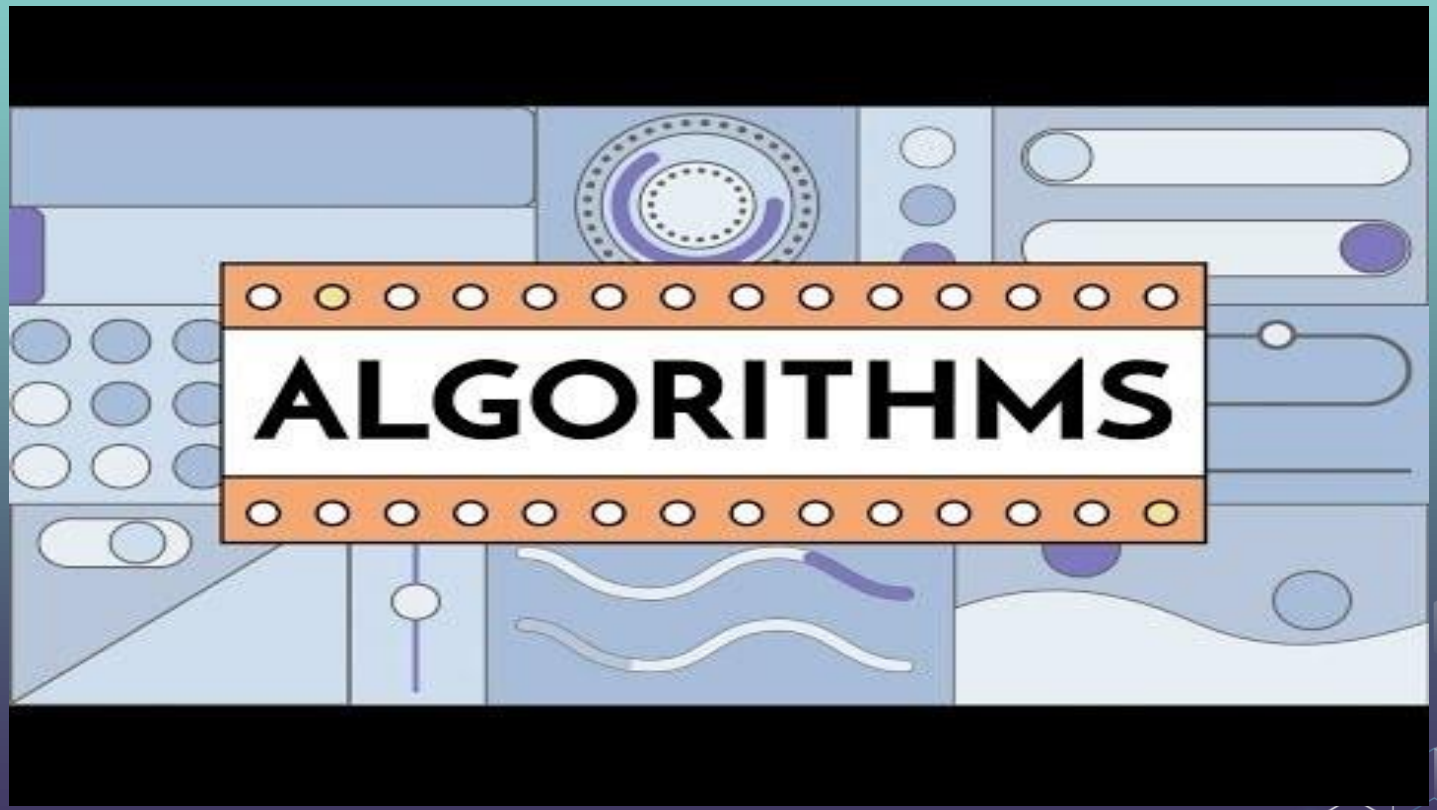
How to collect data?
How to use analytics and algorithms to analyze data?
How to guide firms' business operations using the insights obtained from data?

Questions

What is an algorithm?

Do you know any algorithms?

Questions



What is…

An algorithm?

An algorithm is a finite sequence of well-defined, computer-implementable instructions, typically to solve a class of specific problems or to perform a computation. It is a term used in computer science.

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

What is an algorithm?

Examples of algorithm:

Sorting a sequence of numbers

Finding the shortest path in a network

Autonomous driving

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.

True or False?

This is a course that covers cases and examples.

Mostly False: This is not a typical marketing class that teaches you "soft skills". We focus on hardcore technologies. You will learn different algorithms and data analytics methodologies.

True or False?

I will know AI and Machine learning after taking the class.

Likely False: There are so many AI and ML algorithms nowadays. Even though I will cover several basic ideas in the class, it is impossible for you to become an expert in AI/ML by merely taking this course. You need to learn more.

True or False?

I am already an expert in using Amazon, Google, Taobao, TikTok, Twitter, YouTube etc. There is no need for me to take this course.

False. I am not teaching you how to post on YouTube or how to order products online. We are taking a quantitative approach to understand these online platforms.

True or False?

The course teaches me how to become a successful entrepreneur in the digital age.

Likely False: This course is more technical. You will learn how to collect and analyze data, how to better manage the platform and regulate the industry, but not how to start your own company.

True or False?

The course is very demanding. I will struggle with the course.

False. There will be some workload for you, but I believe this is manageable for most of you.

True or False?

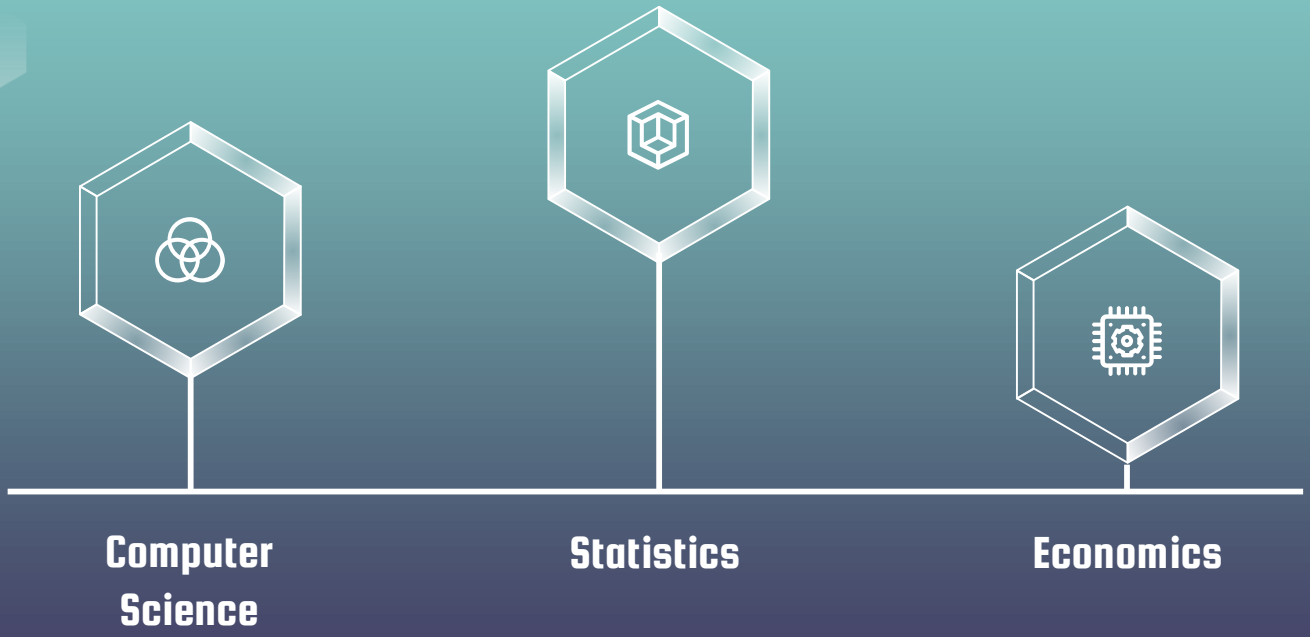
I got a BA degree for my undergraduate studies (我是一名文科生), and I am not good at math. I will fail this course.

False. We are not using calculus or algebra in the class, maybe a bit statistics, but that would also be easy. Perhaps programming would be much more important, but we will start from the very beginning.

This is not a course about. . .

Introduction to Marketing
Web development or Web Design
e-Commerce
Entrepreneurship class - build app/website, become a millionaire!
Preparing a Business Plan, SWOT analysis…

This course adopts methodologies from



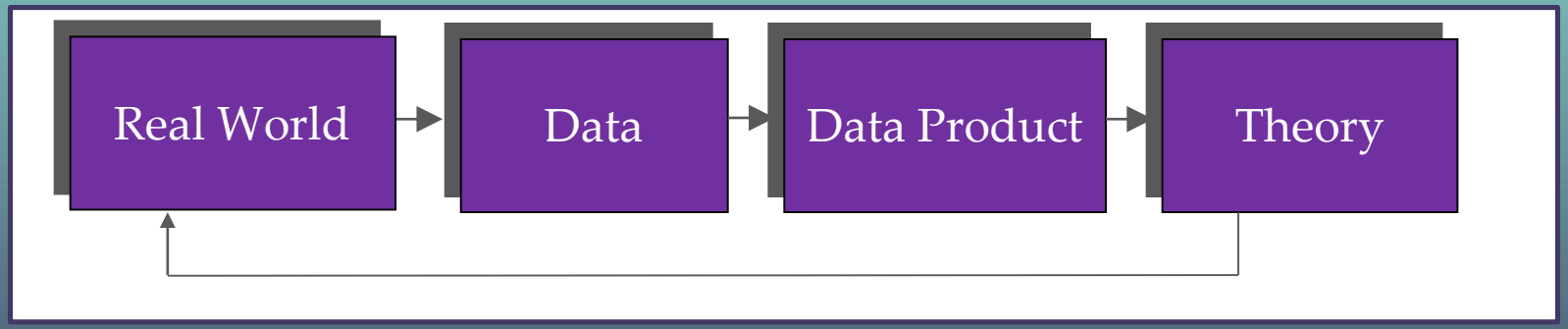
Leveraging your competitive advantage

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

Compared to statisticians and computer scientists, you understand consumers and 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



We emphasize on data, data, and data

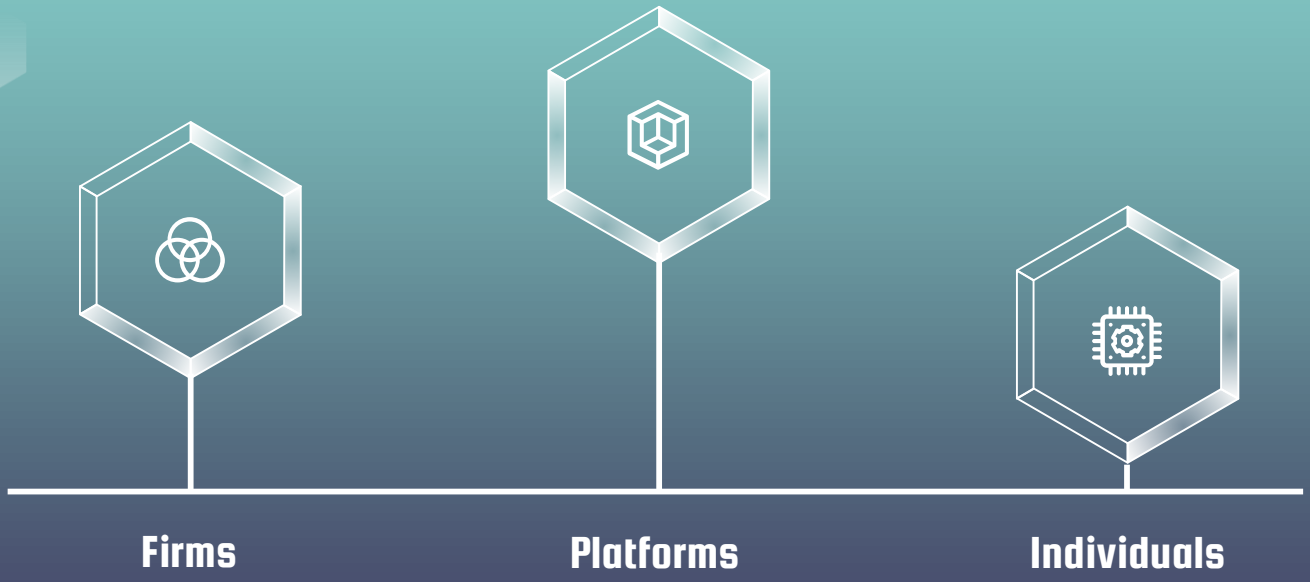
"Talk is cheap, show me the data!"

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.

We solve problems for



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, Associate Professor of Marketing.

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: KKL 836

Our Teaching Assistant

Mindy Jia --- Our alumna, MSc Marketing with distinction

Can be reached via email: mindyjia@hku.hk

My Research

MARKETING SCIENCE
Vol. 34, No. 3, May–June 2015, pp. 331–345
ISSN 0732-2399 (print) | ISSN 1526-548X (online)

http://dx.doi.org/10.1287/mksc.2014.0900

Product and Pricing Decisions in Crowdfunding

Ming Hu, Xi Li, Mengze Shi
Rotman School of Management, University of Toronto, Toronto, Ontario M5S 3E6, Canada
{ming.hu@rotman.utoronto.ca, xi.li13@rotman.utoronto.ca, mshi@rotman.utoronto.ca}

We discuss how crowdfunding, the new online marketplace, differs from traditional online selling platforms.

My Research

Contents lists available at ScienceDirect

IJRM

International Journal of Research in Marketing

journal homepage: www.elsevier.com/locate/ijresmar

Full Length Article

Video mining: Measuring visual information using automatic methods

Xi Li [a], Mengze Shi [b], Xin (Shane) Wang [c,*]

We propose new methodologies for analyzing visual data.

My Research

Article

AMERICAN MARKETING ASSOCIATION

Transparency of Behavior-Based Pricing

Journal of Marketing Research
2020, Vol. 57(1) 78-99

DOI: 10.1177/0022243719881448
journals.sagepub.com/home/mrj

SAGE

Xi Li, Krista J. Li, and Xin (Shane) Wang

We discuss how firms should price discriminate against consumers using consumer data, also known as “杀熟”.

My Research

Reviewing Experts' Restraint from Extremes and Its Impact on Service Providers

PETER NGUYEN
XIN (SHANE) WANG
XI LI
JUNE COTTE

We investigate how different consumers write reviews differently on online platforms (e.g., Yelp, TripAdvisor, and Qunar "去哪儿")

My Research

Audio Mining: The Role of Vocal Tone in Persuasion

XIN (SHANE) WANG
SHIJIE LU
XI LI
MANSUR KHAMITOV
NEIL BENDLE

We propose new methodologies for analyzing audio information.

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: MKTMKT

Do not share the course content with others.

Textbook? No.

Cases? No.

Real Data? Yes.

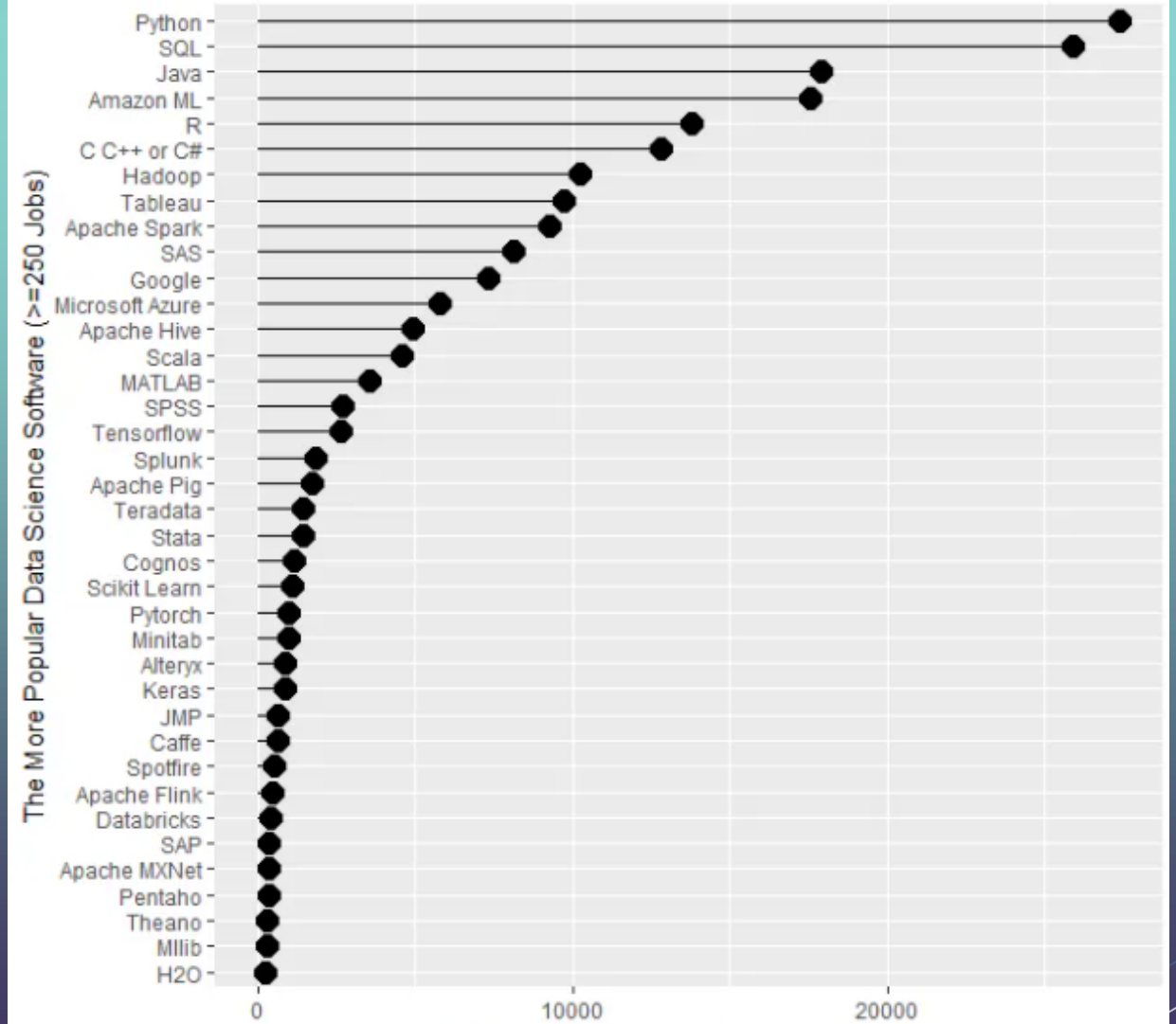
Course Overview

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.


The More Popular Data Science Software (>=250 Jobs)
Python
SQL
Java
Amazon ML
R
C C++ or C#
Hadoop
Tableau
Apache Spark
SAS
Google
Microsoft Azure
Apache Hive
Scala
MATLAB
SPSS
Tensorflow
Splunk
Apache Pig
Teradata
Stata
Cognos
Scikit Learn
Pytorch
Minitab
Alteryx
Keras
JMP
Caffe
Spotfire
Apache Flink
Databricks
SAP
Apache MXNet
Pentaho
Theano
Mllib
H2O
0
10000
20000

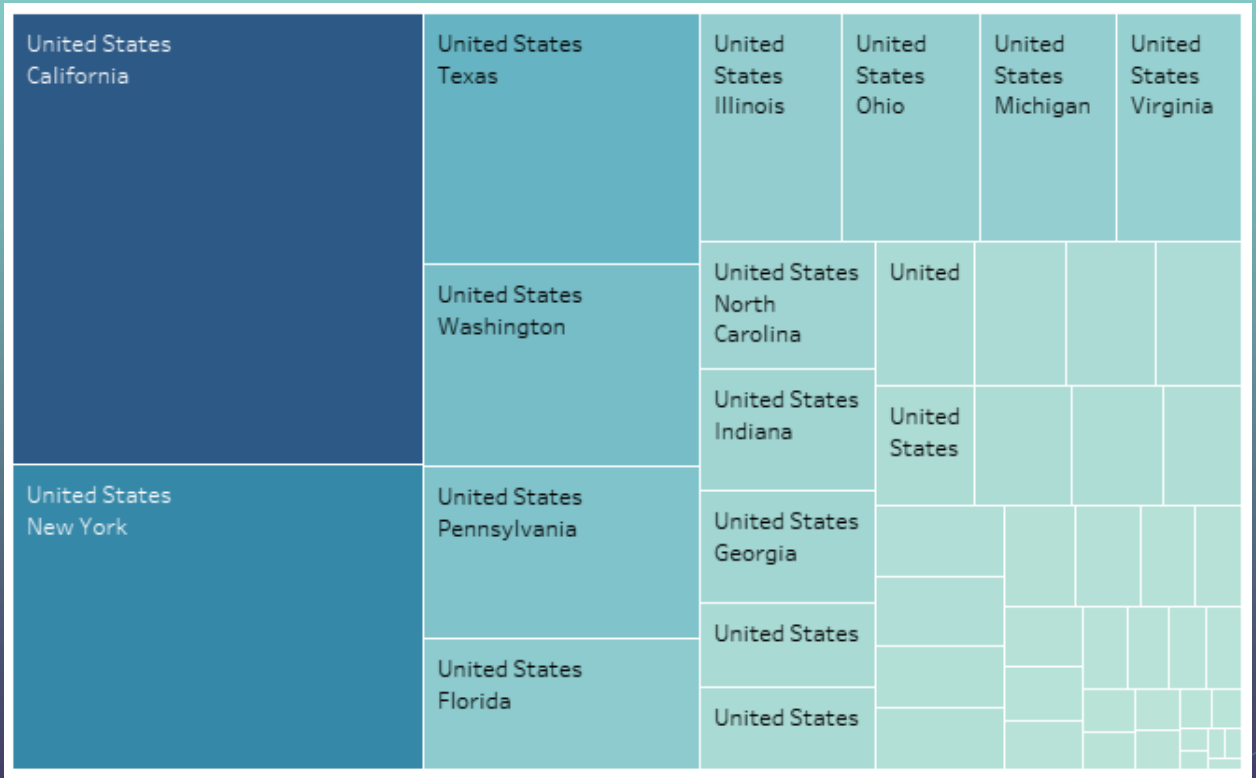
Course Overview

Data visualization with Tableau.

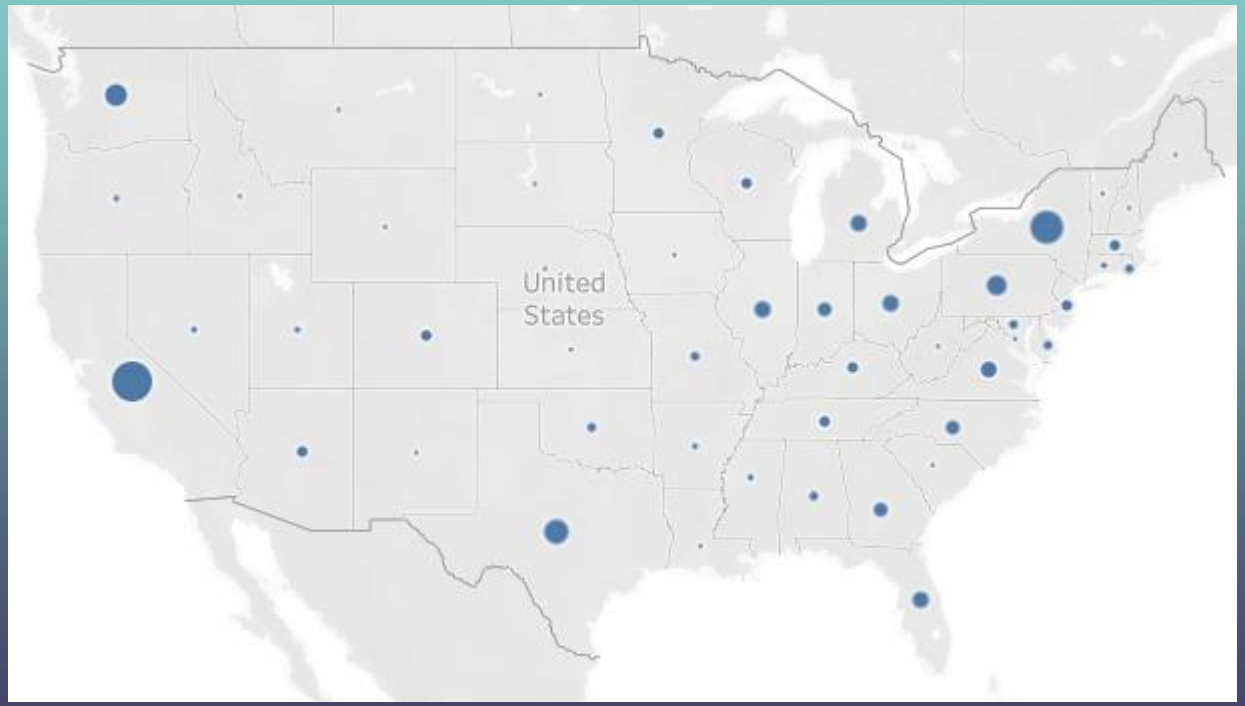
Data visualization gives us a clear idea of what the information means by giving it visual context through maps or graphs.

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, through).

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

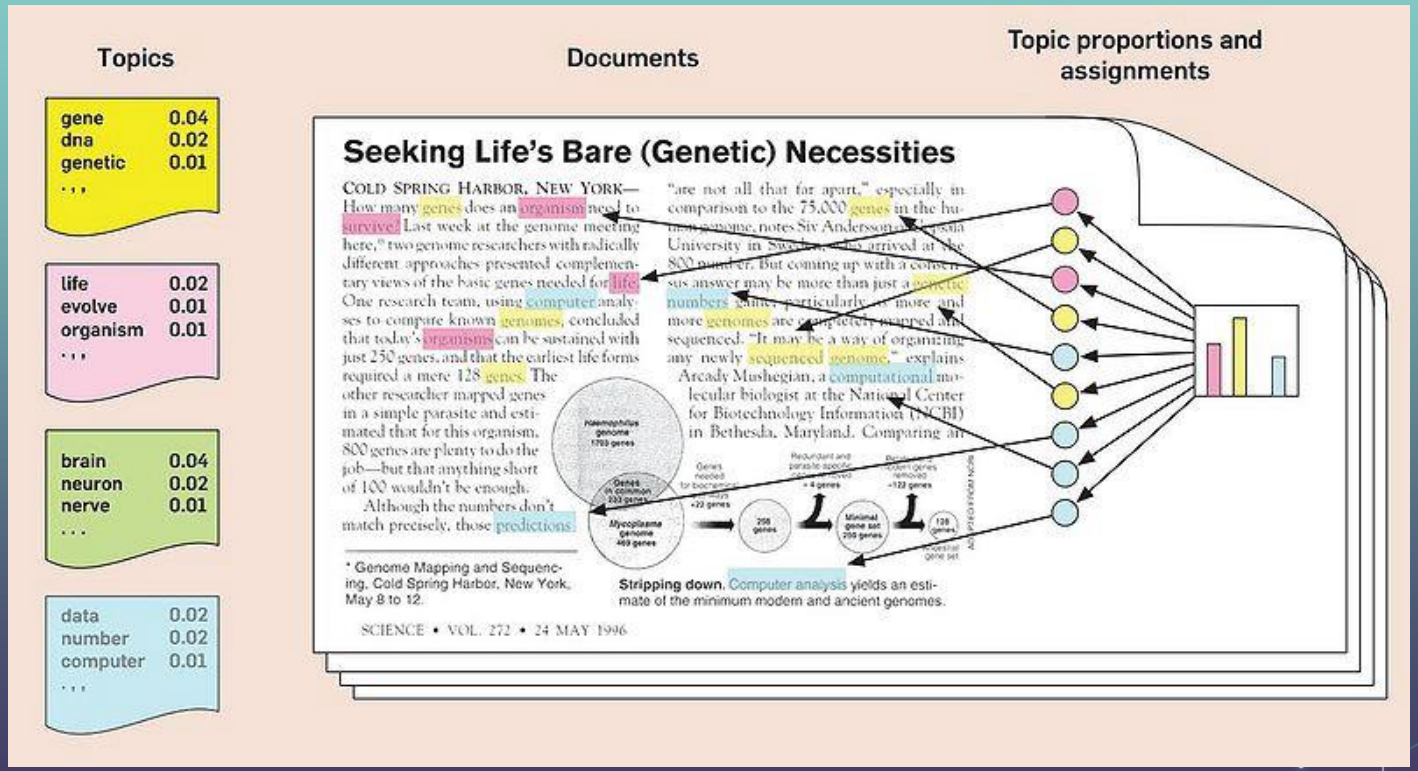
Text Analysis

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



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

40% Group work:

30%: Two data projects, 15% each.
10%: A research presentation.

10% In-class participation:

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

50% Final exam:

Open-book open-notes Multiple choice questions only.

Data Projects

In this course, we are introducing two practice classes 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)
2 data workshops
7 regular lectures

A novel feature of this course

Instead of using cases, we will talk about some real data-analytic project done by professors at top universities (e.g., MIT, Chicago, Harvard) to see the cutting-edge research in digital marketing.

Understand what we can do with data.
Understand how to collect and analyze data, and how to design studies for Internet companies.
Understand what data scientists are working on nowadays.

Mobile ads are more effective in crowded trains

Hyper-C owdedness

Goizueta Business Sc siness, Temple University,

Fox School xm@temple.edu

Stern School se@stern.nyu.edu

Review Ratings Change Sequentially and Temporally

Sequential and Temporal Dynamics of Online Opinion

David Godes
Robert H. Smith School of Business, University of Maryland, College Park, Maryland 20742,
dgodes@rhsmith.umd.edu

José C. Silva
Fuqua School of Business, Duke University, Durham, North Carolina 27708,
josecamoessilva@alum.mit.edu

Schedule

After all groups are formed, I will prepare a list of papers for your presentation. The presentations are spread over lectures, and each class we will have one or more presentations.

How are the papers allocated? --- First come first served.

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”…

Email the TA your group information (group name, your own names and student numbers) before the third lecture.

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。