




Algorithms, Big Data and Online Marketplaces

Welcome to the course!



Survey Questions

1. Do you know anyone who does *not* use the Internet?
 2. How much \$ did you spend on online shopping last year?
 3. Do you rely on online reviews to choose a restaurant, a movie or a product?
 4. Do you trust recommendations made by search engines or e-tailers?
- 

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



Spirit Science


The Internet is changing our lives





Internet creates wealth

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



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?





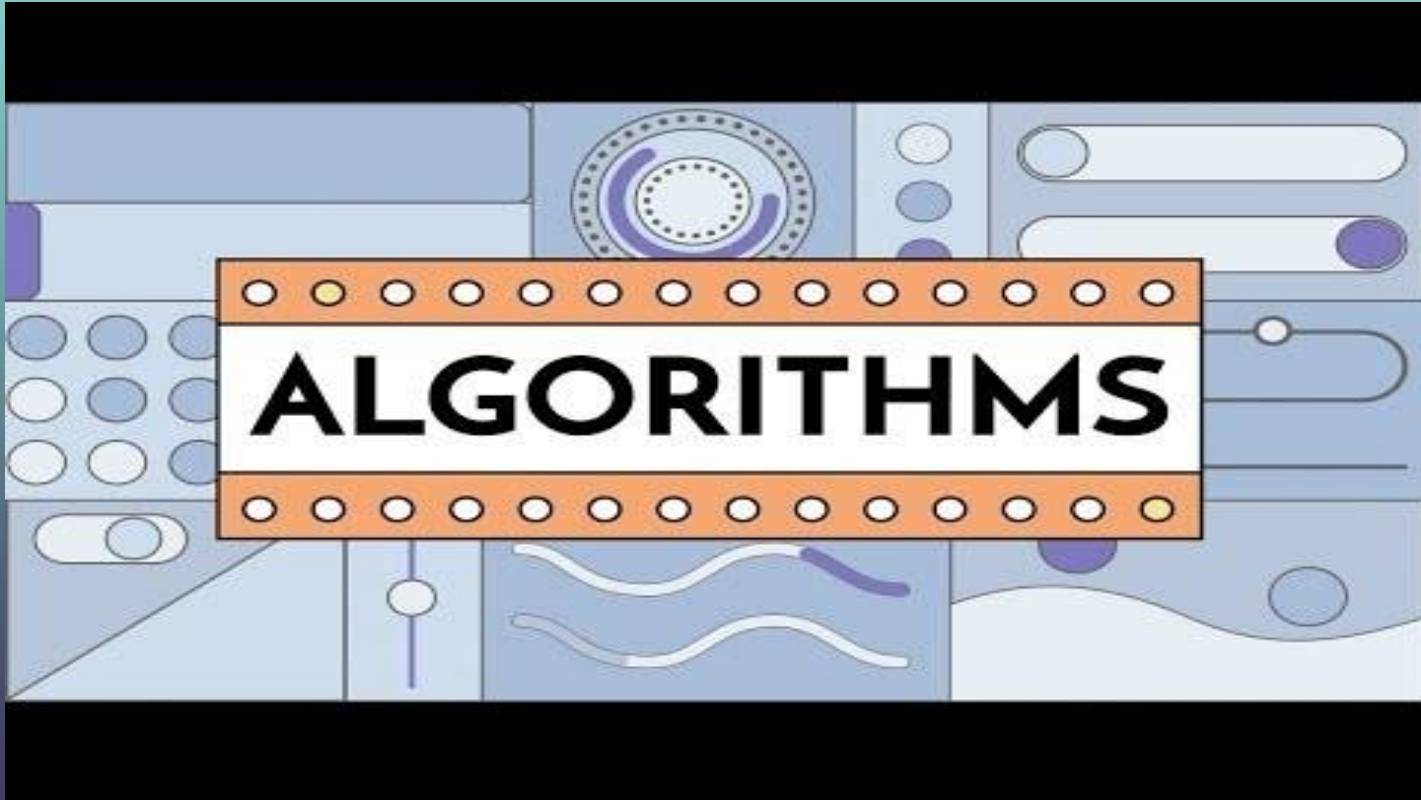
Questions

What is an algorithm?

Do you know any algorithms?



Questions





What is an algorithm?

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.



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



**Computer
Science**



Statistics



Economics




Leveraging your competitive advantage

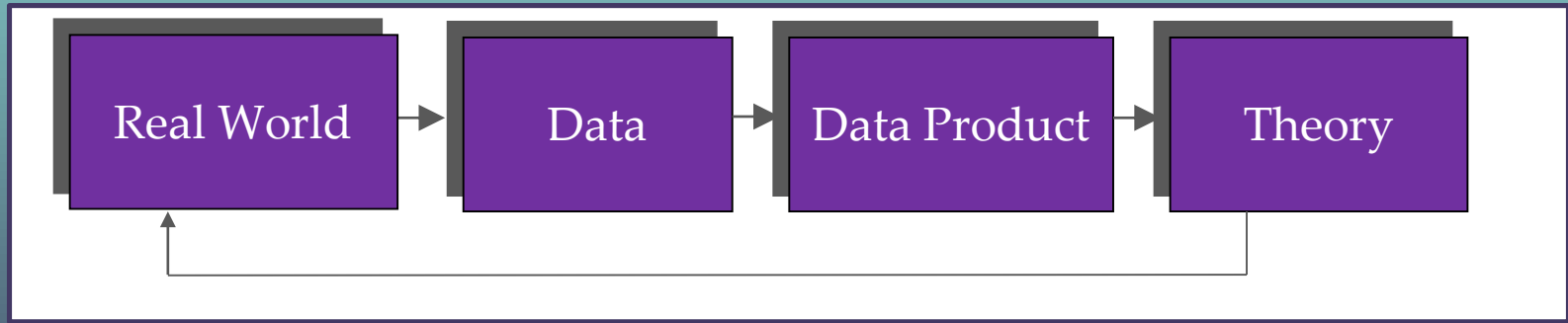
Compared to traditional marketers, you know how to program and how to analyze data.

Compared to statisticians and computer scientists, you understand consumers 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.



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, 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>

© 2015 INFORMS

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



ELSEVIER

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

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

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

© American Marketing Association 2019

Article reuse guidelines:

sagepub.com/journals-permissions

DOI: 10.1177/0022243719881448

journals.sagepub.com/home/mrj



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

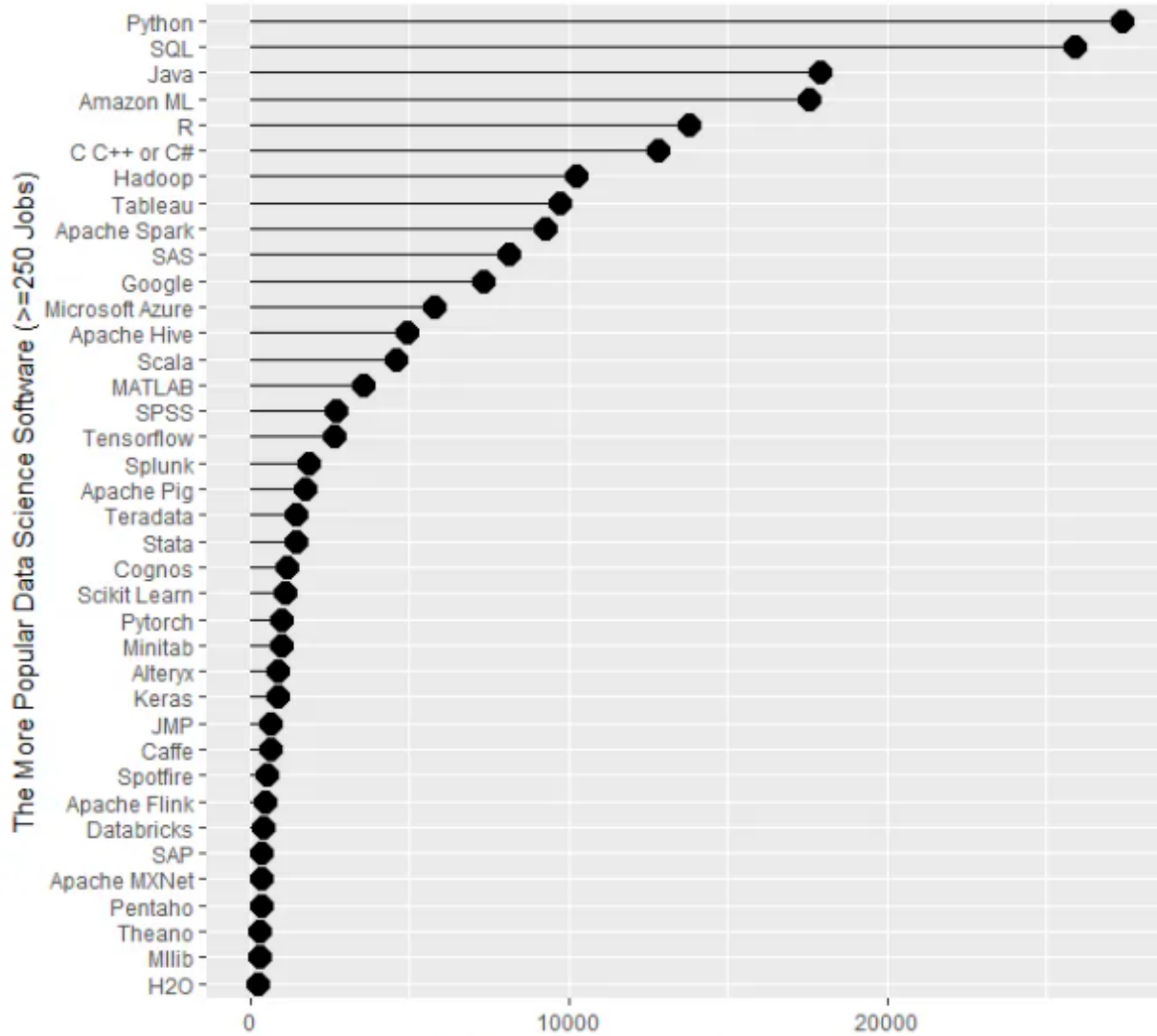
Data analytics with R.

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.








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.

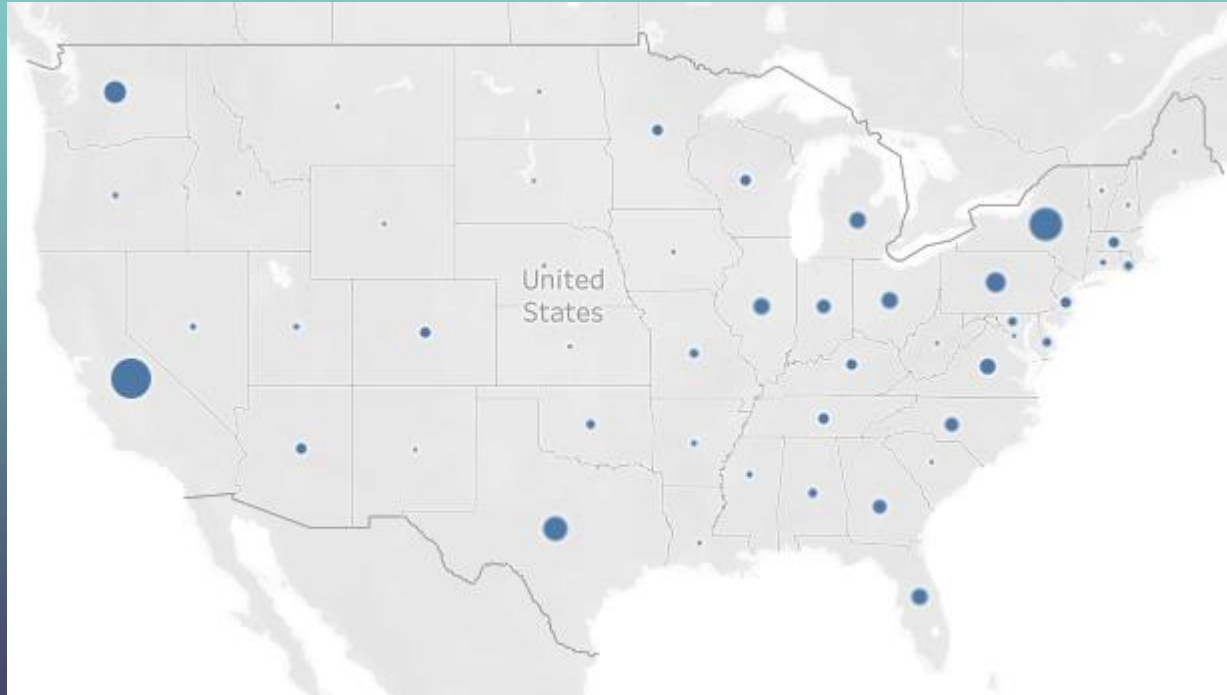
When you want to show your results to others, the best way to use figures --- *“A picture is worth a thousand words”*.



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 HKTVMall, 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, through).

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

Topics

gene	0.04
dna	0.02
genetic	0.01
...	

life	0.02
evolve	0.01
organism	0.01
...	

brain	0.04
neuron	0.02
nerve	0.01
...	

data	0.02
number	0.02
computer	0.01
...	

Documents

Seeking Life's Bare (Genetic) Necessities

COLD SPRING HARBOR, NEW YORK—How many **genes** does an **organism** need to **survive**? Last week at the genome meeting here,* two genome researchers with radically different views of the basic genes needed for **life**. One research team, using **computer** analyses to compare known **genomes**, concluded that today's **organisms** can be sustained with just 250 **genes**, and that the earliest life forms required a mere 128 **genes**. The other researcher mapped genes in a simple parasite and estimated that for this organism, 800 genes are plenty to do the job—but that anything short of 100 wouldn't be enough.

Although the numbers don't match precisely, those **predictions**

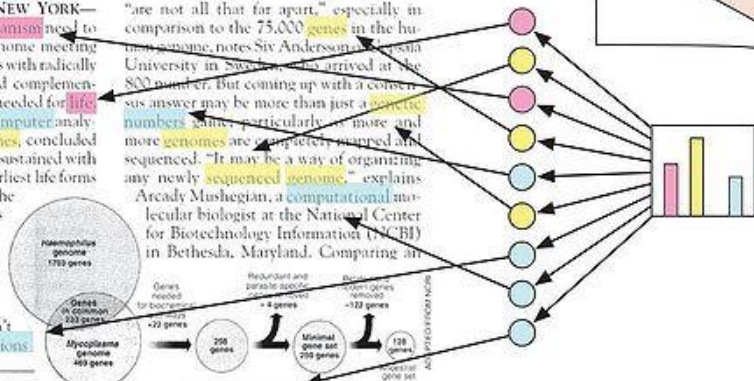
"are not all that far apart," especially in comparison to the 75,000 **genes** in the human genome, notes Siv Andersson, a biologist at Uppsala University in Sweden. "They arrived at the 800 number. But coming up with a consensus answer may be more than just a **genetic** numbers game, particularly if more and more **genomes** are **computer** scanned and sequenced. "It may be a way of organizing any newly **sequenced genome**," explains Arcady Mushegian, a computational molecular biologist at the National Center for Biotechnology Information (NCBI) in Bethesda, Maryland. Comparing an

* Genome Mapping and Sequencing, Cold Spring Harbor, New York, May 8 to 12.

Stripping down. Computer analysis yields an estimate of the minimum modern and ancient genomes.

SCIENCE • VOL. 272 • 24 MAY 1996

Topic proportions and assignments



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:

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

crowdedness

Goizueta Business School

Business, 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,
josecamosilva@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。

