

Business Intelligence Techniques and Applications

Session 1a. Introduction

Renyu (Philip) Zhang

1

About Me



2

2

About Me



3

3

About Me

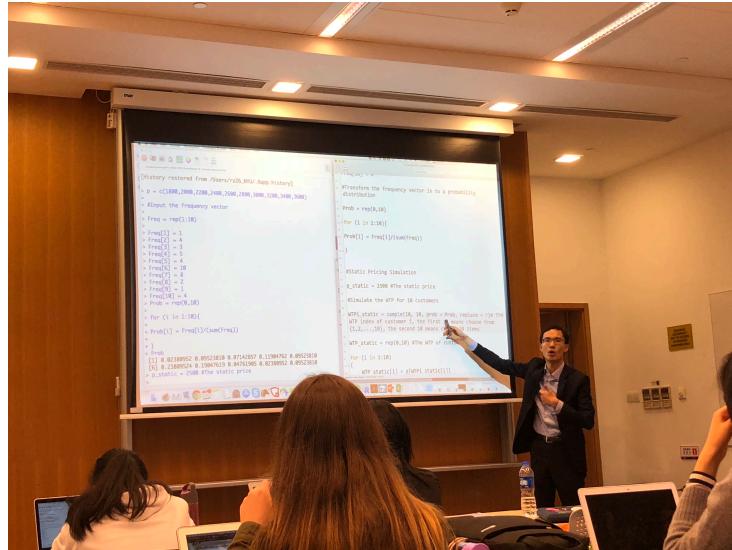


4

4

2

About Me



5

5

About Me



6

6

About Me

A scholar, teacher and practitioner in data science and operations research.

- Visiting Scholar/Associate Professor, CUHK Business School
- Teaching
 - Business Intelligence Techniques and Applications
 - Machine Learning for Business
 - Business Analytics
 - Operations Management
- Research
 - Methods: Data-driven optimization, causal inference, economic modeling
 - Applications: Recommendation, pricing, and advertising strategies of online platforms
- Practice
 - Economist & Tech Lead @ Kwai
 - Data science and analytics in action
 - Evaluation and optimization of recommender systems and advertising platforms

7

7

Who Are You?



8

8

What is Business Analytics?



9

9

What is Business Analytics?

"Big data is like teenage sex; everyone talks about it, no one really knows how to do it, everyone thinks everyone else is doing it, so everyone claims they are doing it."
-Dan Ariely, Duke University



10

10

What is Business Analytics (in this Course)?

- Exciting applications of analytics:
 - Autonomous driving
 - AlphaGo
 - Image and sound recognition
 - Autonomous translation
 - Recommender system
 - Online advertising
 - And many more

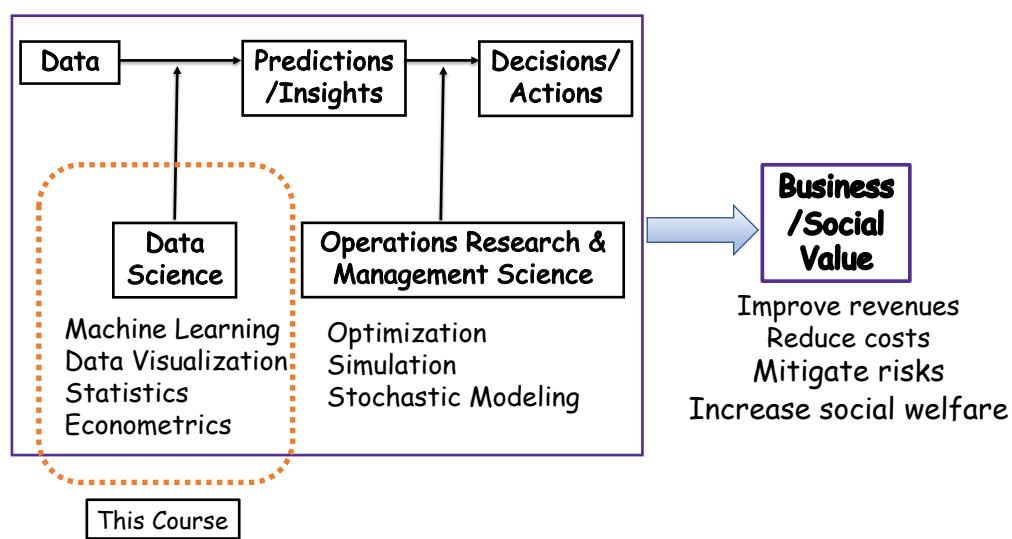
The automated scientific process of transforming **data** into **insights** for making better **decisions** and adding **values** to individuals, companies, and the society.

https://www.youtube.com/watch?v=5adE_cxtSbY

11

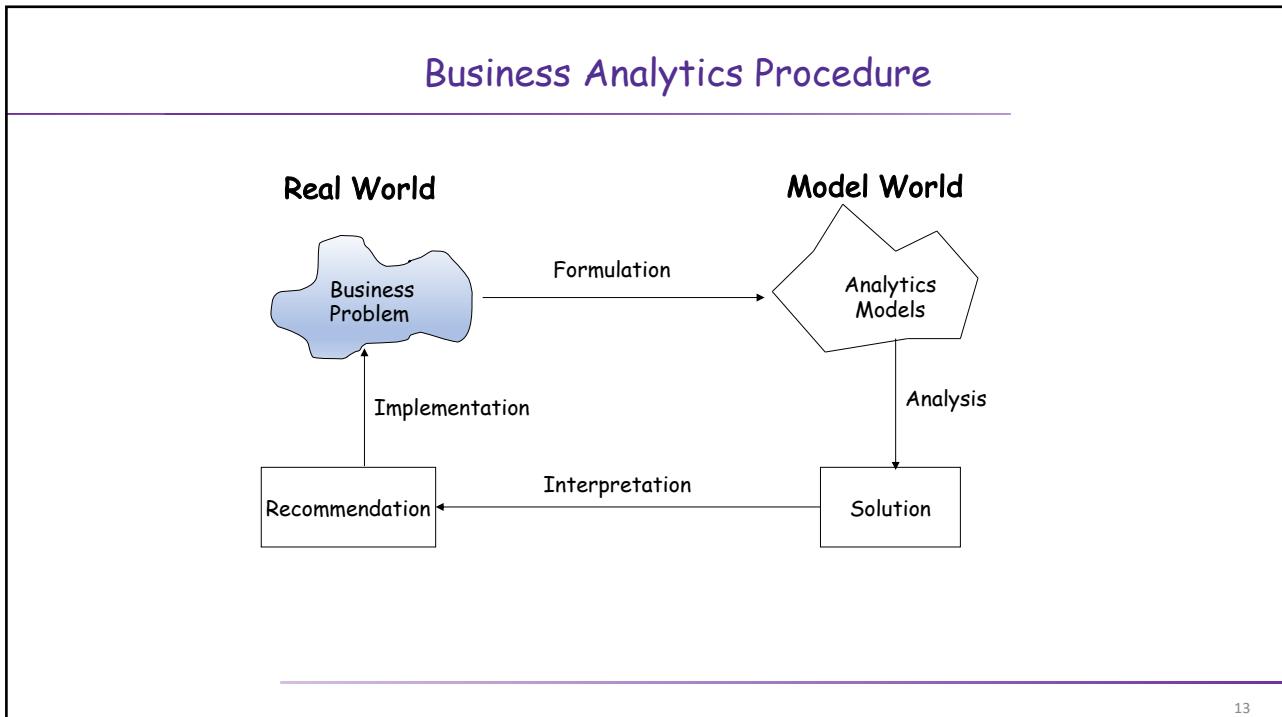
11

Data-Driven Decision Making



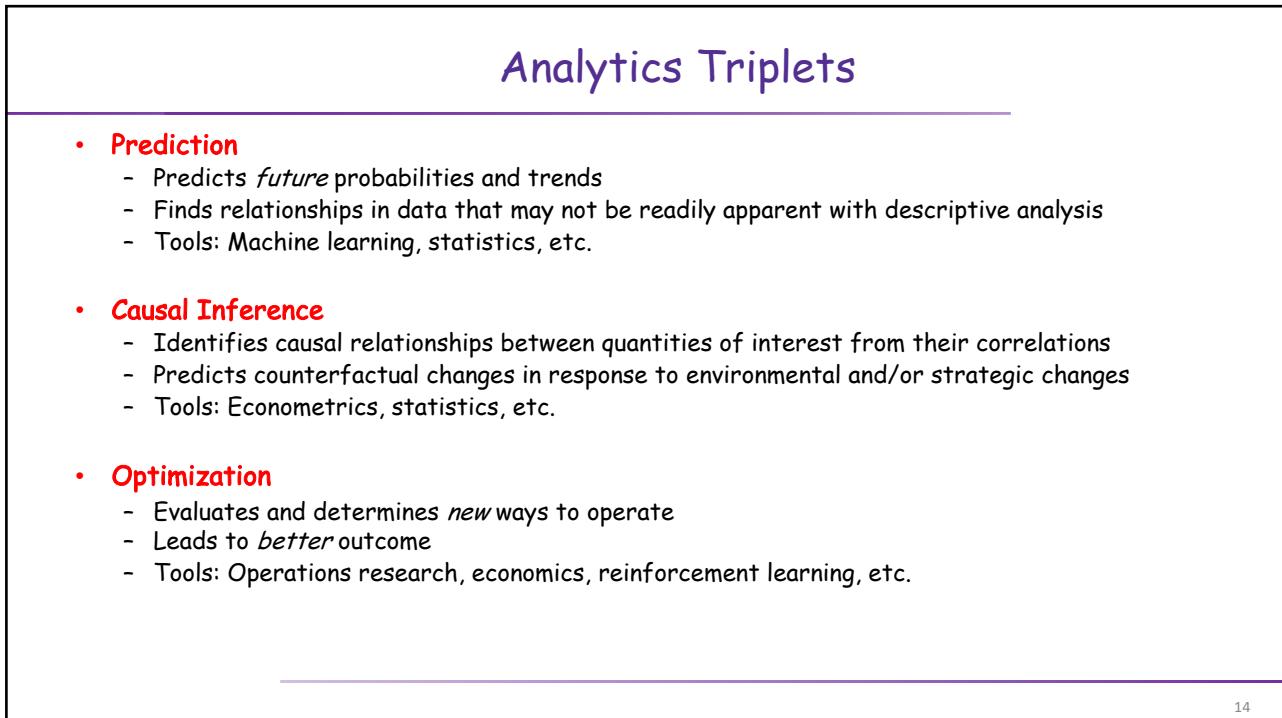
12

12



13

13



14

14

Kwai

- One of the largest short-video sharing and livestreaming platforms.
 - Daily active users: 300 million+
 - Time each user spends on Kwai per day: 120 minutes+
- @ Kwai, you can see different lives and things around the world:
<https://www.bilibili.com/video/BV1YV411r7ns>
- Analytics @ Kwai
 - Making accurate predictions
 - Establishing correct cause and effect relationships
 - Optimizing the recommendation and product strategies



15

15

Prediction @ Kwai



16

16

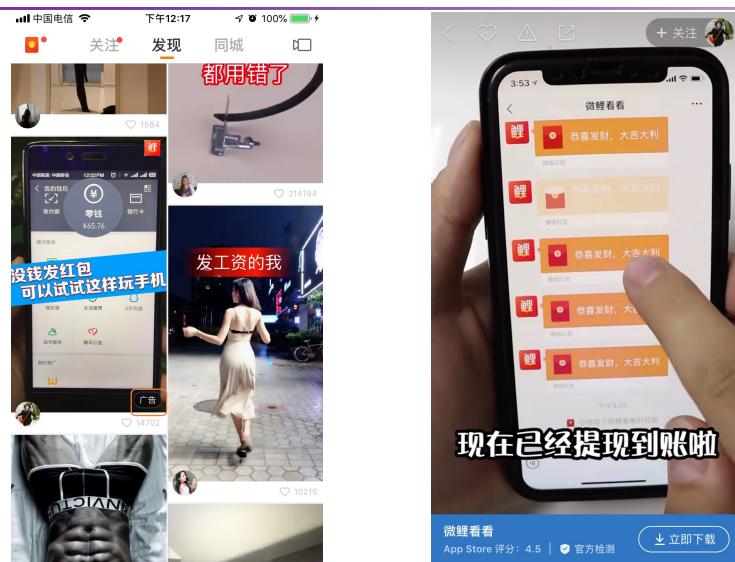
Prediction @ Kwai

- We want to recommend videos you like, so that you will watch them and will stay longer with the platform.
- In order to make good recommendations, we first need to understand who loves what videos most.
- How do we achieve this?
 - Using data about users and videos and machine learning algorithms to predict the probability that a consumer will watch a video and the time s/he will spend on our platform afterwards.
 - Our goal is that for **any user** and **any video**, we can well predict these two quantities.
 - User data: Gender, location, cellphone, watch history, connections, etc.
 - Video data: Producer info. (gender, location, num. of fans, etc.), topic, view num., like num. etc.

17

17

Causal Inference @ Kwai



18

18

Causal Inference @ Kwai

- How would displaying advertisements (adversely) impact user experience?
 - This is important for evaluating the advertising business of Kwai.
- Can we compare the average app time of different users with different advertisement loads?
 - No! More advertisements are displayed to more active users.
- Solution: A/B testing
 - Randomly sample a few subgroups of users with similar feature distributions.
 - Different subgroups are displayed with different ad loads (e.g., 0, 1%, 2%, 3%, 4%, 5%).
 - Comparing the average app time of the subgroups reveals the impact of different ad loads.

19

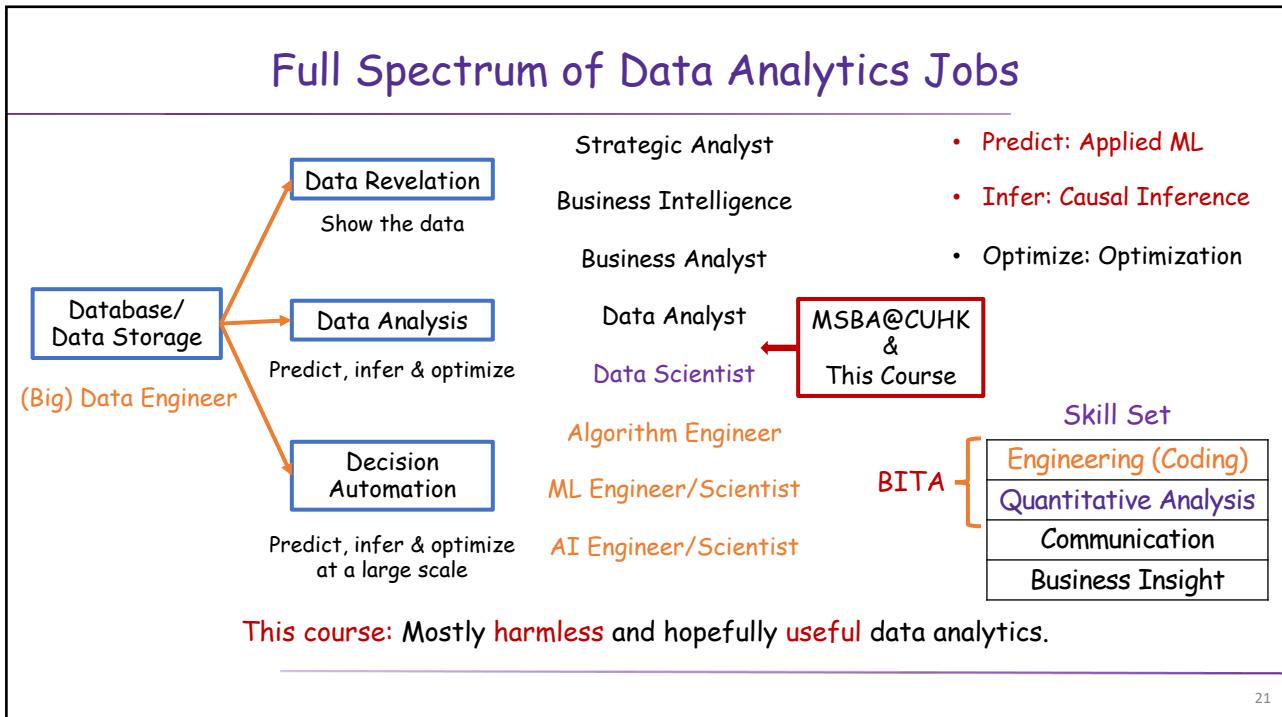
19

Optimization @ Kwai

- How should we allocate consumer traffic to different videos?
 - Intuitively, we should just recommend those who have the most views, likes, and positive comments.
 - How about newly produced videos?
 - Cold start problem.
- Exploration-exploitation tradeoff.
 - Explore new videos vs. Exploit mature ones
 - Current vs. Future
- Solution:
 - Guarantee K displays for each new video.
 - Consumer has a chance of epsilon (a small probability) to be recommended a new (and fit) video.
 - Using optimization techniques to obtain K and epsilon, and A/B testing to validate.

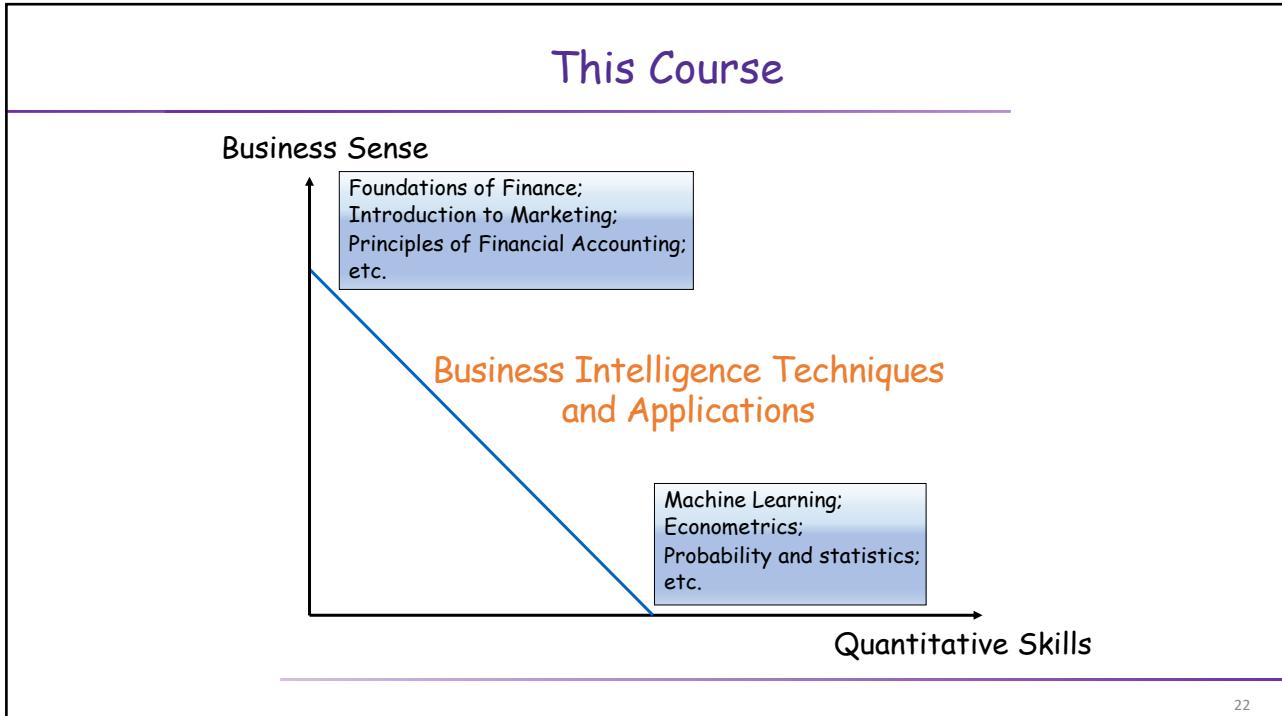
20

20



21

21



22

22

Course Objective

- Our goal: Convince you of the tremendous business (and social) value of analytics and further inspire you to use it in your career and life.
- At the end of this course, hopefully, you will
 - Have an open mind about data;
 - Be ready to be convinced by data and quantitative analysis;
 - Be ready to solve a real problem using data and analytics tools;
 - Be well-prepared to study more advanced analytics courses.
- This is not a math course, not a data science course, not a computer science course, and not even a business course, but an **inter-disciplinary** course that bridges **business applications** and **analytics methodologies**.
- Connection to real business and **job referral** opportunities.

23

23

Course Content

- **Module 0:** Python and data analysis basics
 - Python coding, descriptive analysis, data visualization
- **Module 1:** Prediction with Machine Learning
 - Supervised learning (regression and classification), unsupervised learning (clustering and principal component analysis)
- **Module 2:** Causal inference
 - Potential outcome model, A/B testing, experiment design, experiment analysis, causal inference with observational data

24

24

Course Prerequisites

- No course is required prior to taking Business Intelligence Techniques and Applications, but some knowledge of statistics and coding will be useful.
- Knowledge of basic algebra (including functions such as the quadratic, exponential, and logarithmic) and simple logic is also assumed.
- Not adverse to programming (everything implemented in Python).
- Not adverse to analytical thinking and quantitative analysis in general.

25

25

Learning by Doing

- Lectures and demonstrations (will be recorded)
 - Lecture videos will be posted on Blackboard.
- Extensive cases and data: End-to-end analysis and problem-solving
- In-class discussions
- Weekly problem sets and 2 projects
- Laptop
 - Bring a laptop to every class. **Close your laptop until you are asked to use it.**
 - Install the required applications (Python and Anaconda).
 - Download Jupyter Notebooks (from GitHub) to your laptop before each class.
- Attendance is required.
 - Online attendance is acceptable if you are not in Hong Kong.
 - All electronic devices (except laptops) must be turned off prior to the start of each class meeting.

26

26

Course Materials

- **Blackboard**
 - Syllabus
 - Survey
 - Lecture videos
- **GitHub:**
 - <https://github.com/DSME6756/BA-W2021>
 - Python codes (Jupyter Notebooks) distributed via GitHub
 - Slides and lecture notes
 - Required pre-class readings
 - Additional post-class readings
 - Data and cases
 - Others
- **No required text books**
 - Reference books given in the syllabus and provided at GitHub.
 - Other references will be recommended as we progress to specific topics.

27

27

Course Communications

- **Class Meeting:** Thursday, 9:30AM-12:15PM or 2:30PM-5:15PM, @CYT_410 (you can choose to go to either Section)
- **Office hour:** Thursday, 1:00PM-2:15PM, @CYT_911, or by appointment
- **WeChat group:** Online discussion forum. Please join the group and post your questions, comments, and discussions about this course therein.
- **Instructor contact**
 - Office: CYT_911
 - Email: philipzhang@cuhk.edu.hk
 - Tel: 852-3943-7763
 - WeChat: rphilip_zhang
- **Teaching Assistant:** Huanyu Yin
 - Office hour: By appointment.
 - Email: huanyu@link.cuhk.edu.hk

28

28

Grading

- Class participation, 10%
 - Baseline 6%; lose 2% for each class missed; failing grades for missing 4 or more classes
- Weekly problem sets, $30\% = 5\% * 6$
 - Distributed every Thursday by 6:00PM
 - Due every Thursday at 9:30AM before class
 - 6 problem sets with the highest scores will count
 - Late assignments will NOT be accepted unless approved by the instructor
- Projects, $20\% = 10\% * 2$
 - Project 1: A Kaggle competition, due on Thursday, February 10
 - Project 2: A causal inference case, due on Thursday, March 3
 - For both projects, submit Jupyter Notebooks articulating your analysis and codes
- Final Exam, 40%
 - 6:00PM-9:00PM, March 3
 - Close-book, close-notes, calculators allowed
 - Cellphones, laptops, iPads NOT allowed

29

29

Grading

- Problem sets and projects submitted via Blackboard:
 - Everyone should individually finish and submit his/her own solutions, though discussions with others are allowed.
- Regrading:
 - Submit your requests within 7 calendar days after receiving your grade.
- "Zero-tolerance" policy
 - Any violation of academic integrity is strictly prohibited and will be treated seriously.

30

30

Who Will Find This Course a Nightmare?

- You want an easy A.
- You hate the quantitative/analytical way of thinking and solving problems.
- You hate coding/programming.
- You hate me.

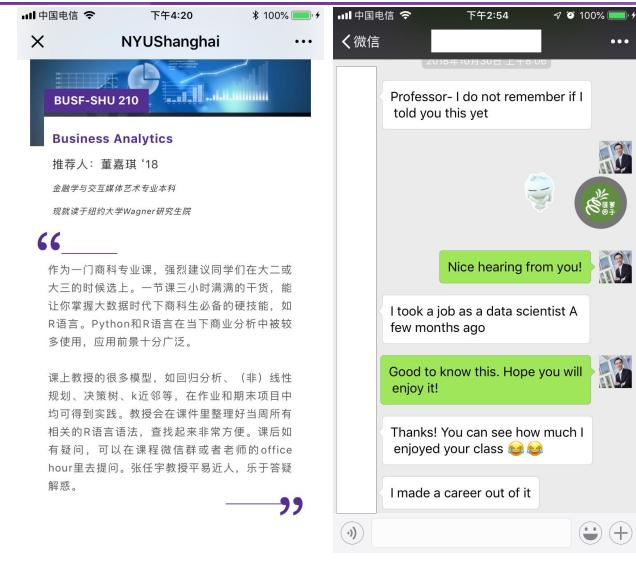
Otherwise, you are very much welcome joining me to enjoy the excitements and challenges of Business Analytics!

31

31



Student Feedback



作为一门商科专业课，强烈建议同学们在大二或大三的时候选上。一节课三小时满满的干货，能让你掌握大数据时代下商科生必备的硬技能，如R语言。Python和R语言在当下商业分析中被较多使用，应用前景十分广泛。

课堂上教授的很多模型，如回归分析、（非）线性规划、决策树、k近邻等，在作业和期末项目中均可得到实践。教授会在课件里整理好当周所有相关的R语言语法，查找起来非常方便。课后如有疑问，可以在课程微信群或者老师的office hour里去提问。张任宇教授平易近人，乐于答疑解惑。

Professor Zhang is very kind and patient with all of his students; obviously very hardworking professor and try his best to bring some of the best and essential materials to be introduced to the students (mainly business students); lay their solid foundational knowledge of analytics; practical hands-on experiences; in a word, help my future career of a business analytics master student and hopefully career as a business analyst/business intelligence analyst/data analyst/analytical consultant/technical consultant.... very much and consolidate my existing knowledge!

32

32