Advanced Business Analytics Practicum

Pragmatic Data Science

Session 1. Introduction

Renyu (Philip) Zhang

1

About Me

2

2

About Me

3

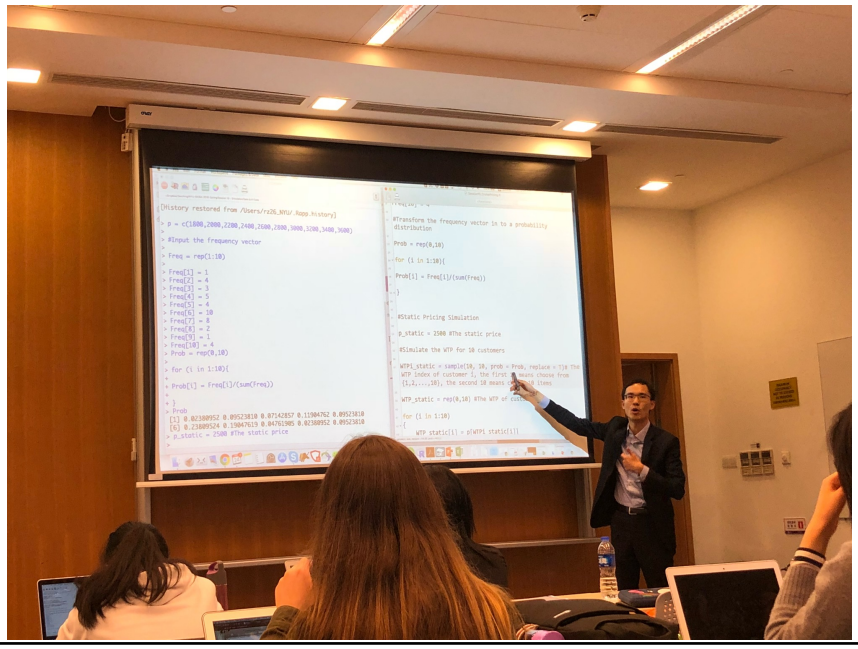
3

About Me

4

4

About Me



5

5

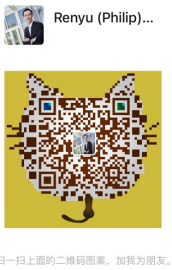
About Me

6

6

About Me

- I study AI and data science to improve business decision making, especially for digitalized online platforms.
- As an economist and Tech Lead of Kuaishou (快手), I build analytics framework to evaluate and optimize the ecosystem of Kuaishou.
- I teach you guys pragmatic data science/AI.

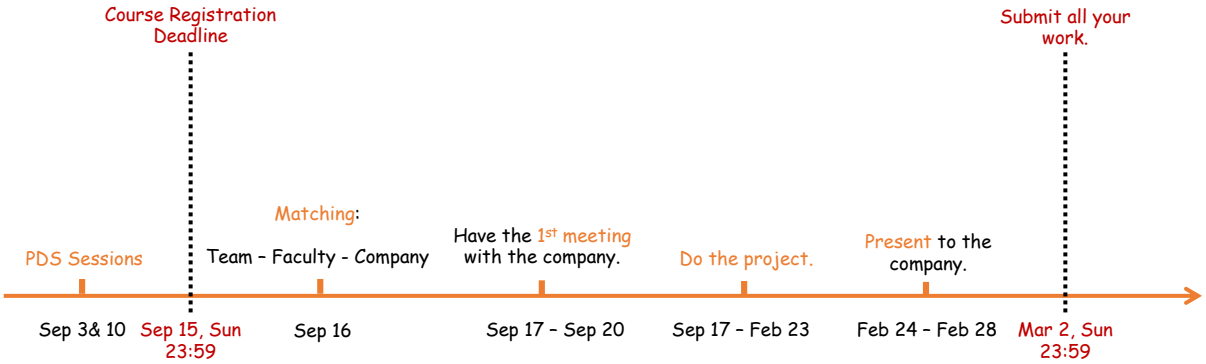


- CUHK Business School, Associate Professor (*with tenure*), since 2022
- NYU Shanghai, Assistant Professor, 2016-2022; Visiting Scholar, since 2022
- Washington University in St. Louis, PhD, 2011-2016
- Peking University, BS, 2007-2011

7

7

Timeline



8

8

Pragmatic Data Science Training

Two mandatory lectures on pragmatic data science:

- 6:30pm-8:30pm, Sept 3 & Sept 10.
- CYT 201 (for registrants) or Zoom (for hesitators)
- Necessary mindset and toolbox for a mostly harmless analytics professional.

9

9

Registration/Drop-off Deadline

Hard Deadline: 11:59pm, Sept 15

- NO add/drop will be allowed after the deadline.
- Overenrollment requests on CUSIS are only available between Dec 2 and Dec 16.
- Anyone who wishes to enroll in this course, please scan the following QR code and submit the registration form, regardless of whether you have registered on CUSIS.

10

10

Matching

All matching will be finished and announced by 11:59pm, Sep 16.

- We will use your industry/job intention and your academic background to do the matching with your teammates and the company.
- We will then assign a faculty mentor to your group.
- We will also assign a group coordinator, who will serve as the point of contact for your group throughout the project.

11

11

First Meeting

Please schedule your first meeting with the matched company in the week of Sep 16.

- As soon as we connect you with your matched faculty mentor and company, each group coordinator please schedule a meeting with them.
- Try to pin down the exact project with the company as soon as possible.
- **Project proposal (with feedback from faculty and practitioner mentors)**: Due at 11:59pm, September 29 (Sunday).

12

12

Do the Project

Work with your groupmates to finish the project until the end of February 2025.

- Submit a weekly project report by 11:59pm every Sunday starting from October 6.
- Submit a midterm report by 11:59pm, December 8 (Sunday).
- The midterm report should have the feedback from your faculty and practitioner mentors.
- Expected workload: 8 hours per week per student.
- Most of the works will be done online, but you can discuss with the company (if it is located within HK SAR) to see whether there are some onsite opportunities/necessities.

13

13

Final Presentation

Please schedule the final presentation in the week of Feb 24.

- The final presentation should be attended by the company and your faculty mentor.
- Submit by 11:59pm, Mar 2, Sunday on Blackboard:
  - Presentation slides deck;
  - Final report (with feedback from both faculty and practitioner mentors);
  - Code/demo/prototype (if applicable).

14

14

Faculty Mentor

The faculty mentor is to help you. It is still your responsibility to carry out the project.

- You can schedule meetings with your faculty mentor if you encounter any technical challenges, but do not over utilize this opportunity.
- Your faculty mentor will also provide feedback on your work.
- If you have any conflict or tension with the company, please also report to your faculty mentor.
- Your faculty mentor will assign the grade to you.
- Feel free to talk to me if you have any questions.

15

15

Practitioner Mentor

The practitioner mentor is the point of contact for the company.

- You can ask your practitioner mentor questions about your project or the company.
- When you get stuck, try to seek for some help from your practitioner mentor first.
- Your practitioner mentor will also provide feedback on your work.

16

16

Course Communications

- No class meeting after Tuesday, September 10
- Office hour: By appointment
- WeChat group: Online discussion forum
- Instructor contact
  - Office: CYT_911
  - Email: philipzhang@cuhk.edu.hk
  - Tel: 852-3943-7763
  - WeChat: rphilip_zhang
- Teaching Assistant: Qinlu Hu
  - Email: qinlu.hu@link.cuhk.edu.hk

17

17

Course Materials

- Syllabus
  - https://docs.google.com/document/d/13-MqYXHFfyGE5Hi-sNbzCC6oagWhjggI1ifUI1nNLOQ/edit?usp=sharing
- Blackboard
  - Proposal, Weekly Report, Final Report, Slides, and Code Submission
- GitHub:
  - https://github.com/rphilipzhang/DOTE6696-24

18

18

Grading

- Quality of your delivered work: 60%
- Final presentation: 20%
- Final report: 20%
- Industry partner feedback: 20%
- You will also need to evaluate each other within each group (to avoid free-riders), whereas I may calibrate the grades of different teams based on the submitted works and comments.

19

19

Industry Partners

Platform-Based

- Kuaishou
- Xiaohongshu
- Tomato of Love (高竞游戏)
- All Saints Music (万声音乐)
- AviGo/JetBay (掌上包机)
- Prime Union (智联汽车)

Finance

- Guotai Junan
- SynFutures
- Arrakis Green FinTech

Consulting

- Bain
- CSPI Rating
- Datality Lab
- GS1 Hong Kong

Manufacturing/Agriculture/Logistics

- Mengniu (蒙牛)
- DCH Auriga (Hong Kong) Limited
- Hong Kong Air Cargo Terminal Limited (Hactl)
- OseanAlpha USV

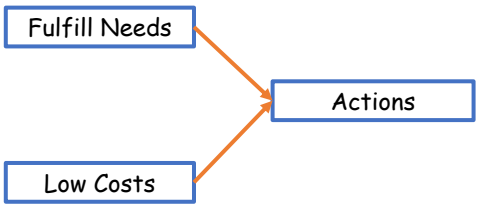
20

20

Fundamental Mindset of Pragmatic Data Science

Can your stakeholder do anything with your results?

- Actionable insights: Your stakeholder can directly improve something following your results.
- Useful knowledge: Your stakeholder can make some long-term strategic shifts based on your results.



A very useful YouTube channel in pragmatic data science: https://www.youtube.com/@pragmaticdata.

21

21

Python and Jupyter

- Python: Very close to English, so not hard to learn.
- Most widely used programming language by data scientists.
  - A huge community with extensive external packages (especially ML & AI).
  - Very easy to find solutions and support when running into problems.
- Jupyter Notebook: A web-based interactive computing environment.
  - Use Anaconda to install Jupyter.
  - Install Python: https://www.python.org/downloads/
  - Install Anaconda: https://docs.anaconda.com/anaconda/install/
  - Alternative: Google Colab https://colab.research.google.com/notebooks/intro.ipynb
- Wisely (or even blindly?) use Google/ChatGPT/Claude/Copilot.
  - https://cuhk-edtech.padlet.org/web/use-of-generative-ai-in-education-h4kuir1lgo42fi0m

22

22

Code Distribution via GitHub

- GitHub (https://github.com/) is a platform for storing code and conducting version control for software development.
  - We use the GitHub for code distribution: https://github.com/rphilipzhang/DOTE6696-24
  - Read this doc for getting started with GitHub: https://docs.github.com/en/get-started/quickstart
- Our GitHub site has only 1 repository (a.k.a. mono-repo).
  - Please remember to pull from this repo to your own computer when necessary.

23

23

Data Base and SQL

- In practice, data are stored in the data base(s) of the firm/organization you work for. You need to pull them out using SQL (Structured Query Language).
  - SQL is the standard language for storing, manipulating and retrieving data in databases.
  - An important technique to gain edge on the job market.
- SQL tutorial: https://www.w3schools.com/sql/, https://www.sqltutorial.org/, http://www.mathcs.emory.edu/~cheung/Courses/377/Others/tutorial.pdf, https://cs.uwaterloo.ca/~tozsu/courses/CS338/lectures/4%20Basic%20SQL.pdf
- SQL questions (and reference solutions) from Leetcode: https://github.com/rphilipzhang/DOTE6696-24

You can learn more in Database and Big Data Management.

24

24

NumPy and Pandas

- Basic libraries in Python to work with arrays (i.e., matrices).
- Foundations of data analysis.
  - Most data analytics tools are based on NumPy and Pandas.
  - https://numpy.org/
  - https://pandas.pydata.org/pandas-docs/stable/index.html
- Important operations:
  - Data loading, extracting, joining, aggregating, etc.
- In this course, you should be reasonably familiar with them (leveraging the help from Google/ChatGPT/Claude/Copilot, etc.).
- Reference book: *Python for Data Analysis*, 2nd Edition, by Wes McKinney
  - Email me if you need a PDF copy of this book.

25

25