

## CS5481 Data Engineering Syllabus

- Course website: Please check the Canvas website often for announcements, course materials, and homework assignments.

We use Jupyterhub for tutorials and course projects. You can get access to the Google CoLab <https://colab.research.google.com/> or the cloud services provided by the department.

- For the homework/projects:
  - Start solving your homework early.
  - The homework is due by **6 pm** on the day noted on the calendar. No late homework will be accepted. Homework will be turned in online via Canvas.
- For the tutorial: tutorials will be mainly about hands-on abilities.
- Grading: HWs= 60% (2 individual assignments with 15% each, 1 group project with 30%), Final= 40% (2 hours).  
For a student to pass the course, at least 30% of the maximum mark of the final must be obtained.
- Team-based learning: Our group project will be based on the university's team-based learning scheme. This is designed to enhance student engagement and peer learning through student-led tutorials, teamwork, and more interactions among peers. Students form groups and collaborate on group projects, including pre-class preparation and in-class group discussions and presentations.
- Instructor: Linqi Song, Associate Professor, CS Department, City University of Hong Kong, email: [linqi.song@cityu.edu.hk](mailto:linqi.song@cityu.edu.hk), office: Yeung Y6415.

TAs:

Ms. Yuxuan Yao (lead TA, [yuxuanyao3-c@my.cityu.edu.hk](mailto:yuxuanyao3-c@my.cityu.edu.hk)),

Mr. Mingyan Liu (lead TA, [mingyaliu8-c@my.cityu.edu.hk](mailto:mingyaliu8-c@my.cityu.edu.hk)),

Mr. Guanzhi Deng (lead TA, [guanzdeng2-c@my.cityu.edu.hk](mailto:guanzdeng2-c@my.cityu.edu.hk)),

Mr. Yunhe Li ([yunheli4-c@my.cityu.edu.hk](mailto:yunheli4-c@my.cityu.edu.hk)),

Mr. Zengyan Liu ([zengyaliu2-c@my.cityu.edu.hk](mailto:zengyaliu2-c@my.cityu.edu.hk)),

Mr. Jilin Cao ([jilincao2-c@my.cityu.edu.hk](mailto:jilincao2-c@my.cityu.edu.hk)),

Mr. Zhuo Han ([zhuohan3-c@my.cityu.edu.hk](mailto:zhuohan3-c@my.cityu.edu.hk)).

- Lecture hours: Wednesday 20:00 - 21:50, Yeung LT-1.
- Tutorials: You only need to attend your session (one of them).  
T61, Wednesday 17:00 - 17:50, Yeung G7510;  
T62, Wednesday 18:00 - 18:50, Yeung G7510;  
T63, Wednesday 19:00 - 19:50, Yeung B7520.
- Schedule: The schedule may be subject to change according to course progress

| Week | Date    | Topic  | HW                       |
|------|---------|--|--------------------------|
| 1    | Sep. 3  | Introduction (data ecosystem)  |                          |
| 2    | Sep. 10 | Data acquisition   | Assign HW1.              |
| 3    | Sep. 17 | Data preprocessing   |                          |
| 4    | Sep. 24 | Data visualization   | Assign group project.    |
| 5    | Oct. 1  | No class (Public holiday)  |                          |
| 6    | Oct. 8  | Advanced topics and recent trends:<br>Large Language Models for Data Engineering | HW 1 due.<br>Assign HW2. |
| 7    | Oct. 15 | Data indexing  |                          |
| 8    | Oct. 22 | Data querying  |                          |
| 9    | Oct. 29 | No class (Public holiday)  |                          |
| 10   | Nov. 5  | Data driven applications (1):<br>Information retrieval and recommendations       | HW2 due.                 |
| 11   | Nov. 12 | Data driven applications (2):<br>Social network analysis and anomaly detection   |                          |
| 12   | Nov. 19 | Data management  |                          |
| 13   | Nov. 26 | Project presentation   | Group project due.       |