

Course Syllabus
SMMG3050: Industrial Data Analytics
2025 Fall

Instructor: Prof. Juan Du
Location and Time: E1-102, 9:00 - 11:50 AM, Friday
Office: W1-506
Office Hour: 3:00pm-5:00pm, Tuesday
Phone: 020-88335989
Email: juandu@hkust-gz.edu.cn
TA: Letian Bai, Rui Wang
Office: W2-5 floor
Office Hour: 3:00pm-5:00pm, Tuesday
Email: lbai799@connect.hkust-gz.edu.cn, rwang834@connect.hkust-gz.edu.cn

Catalog Description (3 Credits)

This course will provide the students with understanding of the fundamentals of using data mining and analytics techniques to transform from data-rich into decision-smart. It focuses on training students with the ability to formulate and solve real industrial problems with the appropriate modeling strategies and scientific principles. This course will include three components: (1) learning fundamental industrial data mining and analytics theory; (2) implementation of industrial data analytics techniques with lab demonstration; and (3) project experience for solving a real-life industrial data analytics problem.

Required Textbook: No

Optional Readings: *Data Mining: Concepts and Techniques*, 3rd edition, written by Jiawei Han, Micheline Kamber, Jian Pei.

Course Outline:

- Introduction to industrial data analytics
- Getting to know your industrial data
- Industrial data preprocessing
- Mining Frequent Patterns, associations, and correlations
- Data analytics and visualization software programming (demo by TA)
- Basic industrial data classification
- Basic industrial data clustering

Learning Outcomes:

At the completion of this course, it is intended that students would be able to:

- Equip with the fundamental knowledge of industrial data analytics concepts, problems, and techniques
- Integrate data analytics techniques with industrial and systems engineering domain knowledge to appropriately formulate problems and facilitate intelligent decision making
- Implement software (e.g., Python) programming skills to transform from data-rich to decision-smart, which will satisfy practical needs and solve application-oriented issues in real industries

- Gain real problem-solving experience through an industrial data analytics project

What you better have known already: basic statistics and basic calculus and linear algebra, and programming. **However, this is not a course of mathematical statistics, advanced calculus, or software course.**

Grades:

Homework (2)	10 % of final grade
Project	25 % of final grade
Presentation	15% of final grade
Exam	50 % of final grade

Assignment of letter grades is based on percentage of points earned out of total points.

Homework: 2 homework assignments will be assigned and due back one or two weeks later. All homework must be submitted by the due date via canvas. Late homework **WILL NOT** be accepted. Collaboration is allowed (and encouraged) on the homework, but you must write the solutions by your own. However, looking at ANY solution other than your own prior to handing in your homework is strictly forbidden.

Exam: You will not be allowed to use your cell phone during exams, even if you are just using it as a clock. Absolutely no make-up exams will be allowed unless a university-approved excuse is provided. When possible, excuses should be provided at least two weeks in advance. If you miss an exam due to a last-minute illness, nothing less than a medical note stating that you were unable to attend will be accepted. Complaints of colds, headaches, sprains, strains, hangovers, late busses, abduction by aliens, etc., will not be accepted. More details regarding final exam will be handed out.

Attendance: Attending class is recommended but not required. If you miss class, you are responsible for any material covered in class that you missed. Being late is disruptive to other students. Please talk with me if something prevents you from arriving at class on time.

Project: The project will ask students to conduct simple data analytics and visualization tasks by using any programming and analytics software except simply asking AI to do the work for you. You may interact with AI agents or tools for help, but you still need to deep think about the project and do it on your own. More information can be referred to HKUST Guidelines and Principles on Using Generative AI for Higher Education (Apr2023). The project guideline for the project will be handed out later. A presentation will take place during the latter part of our class. The final presentation, report and programming files are due on Friday, 11/28/2025.

Presentation: The presentation mainly refers to academic paper presentation. The course provides a paper list, and you may read the paper on the list or find the papers you are interested in and present. The requirement is that the paper

should relate to course content. The Presentation date is scheduled on October 24th, 2025.

Tentative Course Schedule:

Lecture #	Date	Topics	Reference
1	9/5	Introduction to data analytics	Ch. 1
2	9/12	Getting to know your data	Ch. 2
3	9/19	Data preprocessing	Ch. 3
4	9/26	Mining frequent patterns, associations, correlations	Ch. 6
5	10/10	Mining frequent patterns, associations, correlations and industrial research examples	Ch. 6
6	10/17	Basic Classification	Ch. 8, 9
7	10/24	Paper Presentation	
8	10/31	Basic Classification	Ch. 8, 9
9	11/7	Data analytics and visualization by Python	
10	11/14	Classification and Industrial Research Examples	
11	11/21	Basic Clustering	Ch. 10
12	11/28	Project Presentation	
13	12/5	Course overview, outlook, and exam	

Schedule may be adjusted according to students' feedback.