Instructor	Dr. Liu, Jing
Office	WLB516
Phone	3411-7532
Email	jingliu@hkbu.edu.hk
Lecture Hours	14:30 – 17:20 (Sat.)
Venue	DLB303
Consultation	By appointment

Course Objectives:

After taking the course, you should be able to

- Describe the concepts about how big data fits in the organization.
- Identify appropriate data and build statistical models for the data.
- Evaluate model performance and interpret the results.
- Implement data analytics on business problems.

References:

- Provost F. and T. Fawcett (2013). *Data Science for Business*. 1st Ed. O' Reilly Media. **[PF]**
- Jiawei Han, Micheline Kamber, Jian Pei (2012). *Data Mining: Concepts and Techniques*. 3rd Ed. Elsevier. **[HKP]**
- Bernard M. (2016). *Big Data in Practice: How 45 Successful Companies Used Big Data Analytics to Deliver Extraordinary Results*. 1st Ed. Wiley. **[BM]**
- A. C. Müller, S. Guido (2017). *Introduction to Machine Learning with Python*. O'Reilly. **[MG]**
- Wes McKinney (2022). *Python for Data Analysis*. 3rd Ed. O'Reilly Media. **[WM]**
- Joel Grus (2019). *Data Science from Scratch*. 2nd Ed. O'Reilly Media. **[JG]**
- Jake VanderPlas (2022). *Python Data Science Handbook*. 2nd Ed. O'Reilly Media.**[JV]**
- EMC Education Services. (2015). *Data Science and Big Data Analytics: Discovering, Analyzing, Visualizing and Presenting Data*. 1st Edition. Wiley. **[EMC]**

Assessment:

Class Participation	15%	3 in-class quizzes
Written Assignments	30%	3 assignments
Group Project & Presentation	15%	18 minutes each group
Midterm Test	20%	3-hour test
Final Examination	20%	3-hour exam

Preliminary Schedule:

Week	Date	Topic	References
1	6 Sep 2025	Introduction to Big Data Analytics	[PF1-2] [HKP1-2][BM]
2	13 Sep 2025	Python Programming Basics	[WM] [JV2-3]
3	20 Sep 2025	Predictive Modelling: Decision Tree Python Programming – Decision Tree **Assignment 1**	[PF3]
4	27 Oct 2025	Predictive Modelling: Linear Regression Python Programming – Linear Regression **In-class quiz 1**	[MG2.3] [JG14-15] [WM12]
5	4 Oct 2025	Fitting a Model to Data: Objective Functions Python Programming – Logistic Regression **Assignment 2**	[PF4]
6	11 Oct 2025	Overfitting and Its Avoidance Python Programming – SVM & Regularization	[PF5] [MG5.1- 5.2]
7	18 Oct 2025	Midterm Test	
8	25 Oct 2025	Similarity, Neighbors and Clusters Python Programming – KNN & KMeans	[PF6]
9	1 Nov 2025	Data-analytic Thinking: Model Evaluation Python Programming – Metrics **Assignment 3**	[PF7] [MG5.3]
10	8 Nov 2025	Visualizing Model Performance Python Programming – Curves **In-class quiz 2**	[PF8] [MG5.3]
11	15 Nov 2025	Evidence and Probabilities: Naive Bayes Python Programming – Naïve Bayes	[PF9]
12	22 Nov 2025	Association Rules and Itemset Mining Python Programming - Apriori algorithm **In-class quiz 3**	[PF12] [EMC5]
13	29 Nov 2025	Group Project Presentation	