Course Description
In the era of Big Data, the importance of statistics becomes much more significant than before. Tremendous new jobs for statistics such as data science are thus offered nowadays. This course provides fundamental concepts and techniques in the field of data science. Students can take the first step in the world of data science by studying basic concepts in probability and statistics. Some daily-life application of statistics implemented with an open source package Python will also be discussed, e.g. in house pricing, in pricing exchange rates, in classification problems, etc., so as to expose students to practical issues of real-world data.

Topics
1. Probability, Random Variables and probability distributions
2. Point Estimates, Confidence Intervals
3. Hypothesis Testing
4. Linear Regression
5. Logistic Regression models
6. Classification
7. Clustering

Grading Scheme
- Assignments (70%)
- Examination (30%)

[Topics and grading schemes are subject to change as deemed appropriate. Students will receive information and guidelines in class on how they will be assessed for the course.]

Attendance Requirement
Class attendance is expected and required. The minimum attendance required is 70%.

Teaching mode
The course will be conducted via Zoom. We may change to face-to-face session if situation allows.

Instructors
Dr. Chi Wai YU

Dr. Yu is currently working as a lecturer in the Department of Mathematics, HKUST. He obtained his Ph.D. in Statistics at the University of British Columbia. He has more than 10 years of extensive experience in working with statistics and in teaching probability and statistics with different computer software like R, SAS, and Python at undergraduate and postgraduate levels.