# THE UNIVERSITY OF HONG KONG
# HKU BUSINESS SCHOOL

## Big Data Analytics Applied Toward Quantitative Finance
## FINA4359

### GENERAL INFORMATION

Instructor:   Alan Kwan

Email: apkwan@hku.hk
Office: KK Leung 923
Phone: 2858 5614  - please consult via email
Consultation times: By Appointment, otherwise Monday 2-4pm or Thursday 2-4pm.

Tutor: TBC
Pre-requisites: ECON1210 Introductory Microeconomics; *and* IIMT2602 Business applications development *or* COMP1117 Computer programming *or* ENGG1111 Computer programming and applications *or* equivalent;
Co-requisites:
Mutually exclusive:
Course Website: **TBD**
Other important details: I would like to offer concurrently with Module 5 and 6 in the Spring to align with my MFIN teaching.

### COURSE DESCRIPTION

This course provides students a foundation in managing and analyzing large datasets for applications in finance. The first part of the course focuses on building skills – data custodianship and performance computing. Through practice on real-world financial datasets, students will learn methods used to warehouse and retrieve data for high-performance statistical computing. The course then turns to analytical methods with a focus on demonstrating these methods on real-data from various contexts in finance. Methods covered include statistical modeling and inference, machine learning, textual analysis, classification and alternative datasets. Problem sets and projects will be the primary mode of learning. Course learning will be supplemented with exposure to industry speakers from the local financial industry. As for applications, a particular emphasis will be on quantitative trading but course projects will enable a student to pursue his or her own interests.

### COURSE OBJECTIVES

1.   Develop skills in database design, management, and access as would be expected of a first-year investment analyst.

2.   Gain proficiency in programming and performing basic data cleaning, custodianship and data manipulation.

3.   Gain a working understanding of different analytical methods used in finance and where the methods would be appropriate.

4.    Gain fluency for at least one analytical method of the student's choosing through course projects.

……

### FACULTY GOALS

Goal 1: Acquisition and internalization of knowledge of the programme discipline

Goal 2: Application and integration of knowledge

Goal 3: Inculcating professionalism and leadership

Goal 4: Developing global outlook

Goal 5: Mastering communication skills

## COURSE LEARNING OUTCOMES

| Course Learning Outcomes | Aligned Faculty Goals |
|---|---|
| CLO1 – Students will learn to store and access data efficiently using modern database storage methods. | Goal 1, Goal 2 |
| CLO2 – Students will gain an overview of analytical methods used in finance and their typical application, and demonstrate understanding of how to apply the methods through highly-supervised programming assignments. | Goal 1 |
| CLO3 – Students will demonstrate strong fluency in one analytical method of their own choice through course projects. | Goal 3, Goal 4 |
| CLO4 – Students will be encouraged to creatively apply methods or data to solve specific industry problems. | Goal 4 |
| CLO5 – Students will be encouraged to communicate ideas. | Goal 3 |
| … | |

## COURSE TEACHING AND LEARNING ACTIVITIES

| Course Teaching and Learning Activities | Expected contact hour | Study Load (% of study) |
|---|---|---|
| T&L1. - Lectures | 36 | 30% |
| T&L2. – Weekly problem sets. | 48 | 40% |
| T&L3. –Self-learning. | 36 | 30% |
| … | | |
| Total | 120 | 100% |

| Assessment Methods | Brief Description (Optional) | Weight | Aligned Course Learning Outcomes |
|---|---|---|---|
| A1. Problem sets | Students will have multiple problem sets involving high-difficulty programming problems. Each step will enforce a particular skill. Students will be then given a choice of an open-ended problem using example data and be asked to solve the problem. Students can work in groups. | 60% | CLO 1-4 |
| A2. A final project | | 30% | CLO 1-4 |
| A3. General engagement and participation | | 10% | CLO 5 |
| … | | | |
| Total | | 100% | |

## STANDARDS FOR ASSESSMENT

**Course Grade Descriptors**

| | |
|---|---|
| A+, A, A- | The student would can competitively apply analytical methods used in the course independently. |
| B+, B, B- | The student appears to be able to apply analytical methods, but requires guidance. |
| C+, C, C- | The student has a conceptual understanding of methods applied in the course, but could not be expected to apply all methods used in the course. |
| D+, D | The student has shown effort, but a limited understanding of course content. |
| F | The student has not demonstrated effort to understand course content. |

**Assessment Rubrics for Each Assessment** (Please provide us the details in a separate file if the space here is not enough)

For assignments, the main driver of assessment will be accuracy with respect to the answers on which the assignments are based. An "A" quality course assignment looks professional and any discrepancies can be explained carefully. A "B" course assignment is mostly correct, but gaps in understanding remain. A "C" course assignment shows obvious gaps in understanding.

For the final course project, assessment will be based on quality of execution and originality of the investment idea. An A course project will demonstrate thorough understanding of course methods, careful consideration of pitfalls to analysis, and some element of originality. The work will be well communicated and easy to understand.

## COURSE CONTENT AND TENTATIVE TEACHING SCHEDULE

For assignments, the main driver of assessment will be accuracy with respect to the answers on which the assignments are based. An "A" quality course assignment looks professional and any discrepancies can be explained carefully. A "B" course assignment is mostly correct, but gaps in understanding remain. A "C" course assignment shows obvious gaps in understanding.

For the final course project, assessment will be based on quality of execution and originality of the investment idea. An A course project will demonstrate thorough understanding of course methods, careful consideration of pitfalls to analysis, and some element of originality. The work will be well communicated and easy to understand.

I assume a ten week course schedule. I am going to provide assignments that are likely too many for a course of this length, but the number of assignments will be optional according to the course policy described in the section on **Course Policy**.

Programming will be kept on Python and R as a student choice.

1. **Module 1: Working with data**
   a. Lecture 1: Overview of big data in finance, how data is stored and used, introduction to data warehousing
   b. Lecture 2: writing SQL queries, when not to use SQL.
2. **Module 2: Working with data – continued)**
   a. Lecture 1: Manipulating data retrieved from a database. Coding for performance.
   b. Lecture 2: Manipulating data continued. Data visualization. Parallel computing architectures
3. **Module 3: Working with financial data**
   a. Lecture 1: Datasets used in finance
   b. Lecture 2: Cleaning and programming data
4. **Module 4: Basic analytical tools in finance**
   a. Lecture 1: Regression methods
   b. Lecture 2: Big data statistics
   c. Assignments
      i. Assignment related to default probability
      ii. Assignment related to linear regression
5. **Module 6: Quantitative trading**
   a. Lecture 1: Traditional trading strategies (value, momentum)
   b. Lecture 2: Performance back-testing
   c. Lecture 3:
6. **Module 5: Alternative data**
   a. Lecture 1: Alternative data
      i. Applications to quantitative trading strategies
   b. Lecture 2: APIs and webscraping
7. **Module 7:**
   a. Lecture 1: Basic machine learning - regression trees, cluster analysis
   b. Lecture 2: Reserved
   c. Required assignment:
      i. Course project involving heavy use of one prior method to solve a problem that is standard in finance. I will provide project goals. Students will see the project list from the first week of class.

Assignments will vary year by year.

## REQUIRED/RECOMMENDED READINGS & ONLINE MATERIALS (e.g. journals, textbooks, website addresses etc.)

## MEANS/PROCESSES FOR STUDENT FEEDBACK ON COURSE

○ conducting mid-term survey in additional to SETL around the end of the semester

○ Online response via Moodle site

○   Others: _____ (please specify)

**COURSE POLICY** (e.g. plagiarism, academic honesty, attendance, etc.)

Group work encouraged. Students will describe their contributions to each with a signature. Attendance is encouraged and will factor into partial evaluation. No testing, and therefore no plagiarism.

**ADDITIONAL COURSE INFORMATION** (e.g. e-learning platforms & materials, penalty for late assignments, etc.)

This course adapts MFIN 7035 – Big Data in Finance – to the undergraduate level with some small part splicing in teaching on Quantitative Trading, as allowed.