Capital University of Economics and Business
Overseas Chinese College
Course Syllabus

Year and Semester	2025 Fall					
Course Name	Python Crawler and Data Analysis					
Course Code	MIS345					
Course Type	☐ General Education (Required) ☐ General Education (Elective) ☐ Basic Disciplinary Course ☐ Professional Course (Required) ☐Professional Course (Elective) ☑ Professional Course (Expanded) ☐Professional Course (Advanced)					
Course Credits	4					
Course Hours	Total Class Hours	64	Lecture Hours	32	Experiment (Computer) Hours	32
Applicable object	☐ Freshman ☐ Sophomore √ Junior ☐Senior					
	☐ Business Administration (Accounting)					
	√ Information Management and Information Systems (Finance)					
Prerequisites	None					
Instructor	Jingning Li					
Contact Information	Office: C217					
	Tele: (010)83951082					
	Email: lijingning@cueb.edu.cn					
Office Hour	MW: 9:55-11:30; TH: 13:30-15:05					
Learning Centre	W: 15:25-17:00, 18:00-20:00 (online)					
Grade/Section	2023IT					
Course Time/Place	2023IT: M 13:30-15:05, TH: 9:55-11:30 / B307					
Textbook	Wes Mckinny, *Python for Data Analysis, 1st Edition . O'Reilly Media,Inc*, ISBN 978-1-449-31979-3					

Reference Book

1. Pyan Mitchell,*Web Scraping with Python,1st Edition. O'Reilly Media,Inc*,, ISBN 978-1-491-91027-6.
2. Gabor Laszlo Hajba,*Website Scraping with Python,1st Edition. Apress* ,ISBN 978-1-4842-3925-4.

Course Description

This course consists of "Python Web Crawler" and "Python data analysis". It completely explains the main technologies necessary for data crawling and data analysis with Python, including data crawling, data storage, data processing, data analysis, data visualization, etc. Then through a number of practical cases, the technical knowledge is applied to meet the actual needs of business. This course is helpful to cultivate students' Computational Thinking, data thinking and practical ability to solve business problems by programming.

Student Learning Objectives

On successful completion of this course, candidates should be able to:

Knowledge	◆ explain the principle of Python crawler ◆ recognize the process of Python data analysis ◆ master parsing HTML with Python ◆ master methods of Python crawler ◆ skillfully operate skills of Numpy and Pandas
Capability	◆ crawl the required network information by using Python ◆ master the skills of data loading, storage and analysis by using python ◆ make chart and visualization according to requirement by using python ◆ analyze and solve business problems by using python
Mindset	◆ cultivate students' sensitivity to data ◆ establishing students' consciousness of quantitative analysis ◆ train students' thinking of speaking with data ◆ enhance students' professional quality of data analysis

Website Source

1.https://www.bilibili.com/video/BV1Ax41177LJ?from=search&seid=2144148978818625742
2.https://www.bilibili.com/video/BV1QW411a71e?from=search&seid=2144148978818625742

Teaching Methods

This course includes skill demonstration, project practice, homework and classroom test. In the last two weeks, each student will be provided with personalized data to test their ability to understand and apply Python knowledge.

This course adopts the flipped classroom teaching mode, and provides detailed Python operation handouts in advance. Students are required to complete the preview and homework before class, assess and score in class, finish the project cases independently after class, and obtain the final results by means of speech competitions.

Grade Criterion

Component	Weight	Description
Final Exam	20%	A cumulative final examination will be given based on all of the contents of the class. The exam paper may be composed of multiple-choice questions, short answer questions, essay questions, problems, and preparation of financial statements. Students should rely primarily on homework assignments to give them a sense of what they may see for material on exams.
Mid-Term Test	20%	A cumulative midterm test will be given based on all of the contents that have been taught in class. The test paper may be mainly composed of multiple-choice questions and it should be completed within 15 minutes in class.
Homework	15%	Most of the assigned homework is taken from the Exercises in the textbook. Assignments will be collected at the clearly stated date. Late assignments will not be accepted. The graded assignments will be kept by the tutor for reference and won't be returned to students.

Quizzes	15%	There will be at least 2 quizzes during the semester. Quizzes may or may not be announced in advance. It may also be used as a way to check the attendance. Quizzes will test your knowledge of both concepts and the application of those concepts.
Presentation	10%	The students will be divided into several groups to prepare a presentation. Each student is required to be involved in the presentation. The topics can be selected from the textbook or lectures. Each group need to finish a PPT related to the topic which is given and hand in the related resources to the teacher before the presentation.
Participation	10%	Individuals will be asked to participate individually in a question and answer at least 5 times during the semester. The performances should be counted in their participation.
Attendance	10%	Refer to attendance policy listed below
Total	100%	

Detailed Grade Computation

	Before Midterm	After Midterm
Attendance	5%	5%
Participation	5%	5%
Homework	5%	10%
Quizzes	5%	10%
Presentation		10%
Mid-Term Test	20%	
Final exam		20%
Total	40%	60%

Assessment of Student Performance

☞ ***Self-Study and Reading ability Practice***

Instructor will give out the chapters or the reference books to read and use class hours to have discussion; students should be able to show a proactive attitude and ability for self-study and reading. Knowledge and oral English will be elements of homework or presentation score.

☞ ***Homework***

Students should finish their homework by themselves. Copying from others will be treated as cheating and the homework scores will be lowered. Students should hand in all assignments on time. Late assignments will be accepted at the discretion of the instructor (i.e., when the student was ill or had an excused absence). Late assignments without reasonable proof will be reduced in score by 50%.

☞ ***Attendance***

Because the course covers a great deal of material, attending every class session is very important for performing well.

- Being late for 15 minutes or more is considered an absence.
- Five hours or above of unexcused absences will result in the lower level of the final grade by one grade band (e.g. from C – to D +). Any excused absence must be discussed directly with the teacher.

- Absence which is more than 1/3 of the total teaching hours will cause an F (a failing grade) directly. but students are welcome to continue attending classes.
- An incomplete grade (I) will be considered in case of medical or family emergencies.

☞ ***Participation***

- Students should participate in classes actively. Half of participation grade is determined by their presentation in class. They are encouraged to ask questions relevant to the subject and express their own opinions. Every student should respect the ideas, opinions, and questions of their classmates.
- Students should also use office hours to ask questions or talk with the instructor for good communication and effective learning.
- Frequent visiting the instructor and chatting in English during office hours is highly recommended.
- Any misbehavior and non-class related activities in class will result in the lower level of the participation grade, including ringing cell phones.
- All above behaviors will be solely evaluated by the instructor for scoring.

☞ ***Textbook***

Students must bring the textbook to class.

Topical Course Outline (original)

Week	Date	Topics	Homework
1	Sep.8	Introducing Syllabus Introduce the textbook and 2 projects **Review Power BI – Crawler** ●·Power BI: √ Introduce Power BI's 4 parts: Power Query, Power Pivot, Power View, Power Map √ Introduce Power BI's 6 steps: Get data, manage data, build tables' relationship, create metric-values, create visual reports, add into PPT √ Introduce Power BI's basic operation: workbook, worksheet √ Import data and crawling data from the website (Chapter 1-4)	Setup a group with 3-4 classmates
	Sep.11	●Power-BI: √ Manage data: Power Query √ Manage rows and columns (Chapter 5-6)	
2	Sep.15	**Introduce Python Crawler - BeautifulSoup** **Method 1**: Tag/Element selector	
	Sep.18	Task 1	
3	Sep.22	Introduce Python Crawler **Method 2**: Tag/Element hunter – find() and find_all()	
	Sep.25	Task 2	
4	Sep.29	Introduce Python Crawler	

		Method 3: CSS selector - select()	
	Oct.2	National Day Holiday	
5	Oct.6	National Day Holiday	
	Oct.9	Task 3	
6	Oct.13	**Presentation – AI Crawler**	
	Oct.16	**Presentation – AI Crawler**	
7	Oct.20	**Review Power BI – Data Analysis** ●Power-BI: √ Create data analysis models: DAX (Chapter 7) ●Power-BI: √ Create a data visualization report √ Edit the data visualization report √ Beautify the data visualization report (Chapter 8-10)	
	Oct.23	**Introduce Python – Data Analysis** ● Review Python ● Introduce NumPy ● NumPy Case Study	Case Study: Analysis of the returns on bank financial products
8	Oct.27	● Review NumPy ● Introduce Pandas	
	Oct.30	● Review Pandas ● Continue Introducing Pandas	
9	Nov.3	● Continue Introducing Pandas ● Review Pandas - DataFrame	
	Nov.6	● Continue Introducing Pandas ● Pandas Case Study	Case Study: Stock analysis and trend
10	Nov.10	● Introduce SciPy	
	Nov.13	● Continue Introducing SciPy ● Introduce Matplotlib (18 basic drawings)	
11	Nov.17	● Continue Introducing Matplotlib (10 basic diagrams)	radar chart
	Nov.20	● Continue Introducing Matplotlib (5 complex type drawings)	
12	Nov.24	● Continue Introducing Matplotlib (Comprehensive Diverse Chart) ● Pandas Direct drawing	
	Nov.27	● Introduce Seaborn: Relationship mapping（scatter plot）	
13	Dec.1	● Continue Introducing Seaborn：Relationship mapping（line chart）	
	Dec.4	● Continue Introducing Seaborn：Categorized plotting ● Continue Introducing Seaborn：Distributed plotting	
14	Dec.8	● Continue Introducing Seaborn：Regression plot	

		● Continue Introducing Seaborn：Matrix drawing	
	Dec.11	● Continue Introducing Seaborn：Advanced drawing	
15	Dec.15	● Introduce PyEcharts	
	Dec.18	● Introduce PyEcharts	
16	Dec.22	**Presentation – AI Data Analysis**	
	Dec.25	**Presentation – AI Data Analysis**	
17	Dec.29	Review	
	Jan.1	Q&A Time	**Submit personal report**

Note: *Some chapters or sections may leave for self-study, this is the students' duty to learn and understand, they may also be included in the quizzes or exams.A review in Chinese may be held during L.C. and O.H. in the semester.*

Teacher's Office Hour

- The instructor's office hour is shown in the front of the office door.
- Students are suggested to use the instructor's office hour and learning center to ask questions or talk with the instructor once at least per week for good communication and effective learning, which is recorded in the students' participation.
- The time can be scheduled by instructors or students, or both.

Cheating and Plagiarism

Cheating is not tolerated. Any student caught cheating on a quiz; test or exam will be given a mark of zero (0) for the particular work. At the beginning of the semester the definition of plagiarism will be carefully explained, when any thoughts or writings of another person are used, they must be clearly identified (usually one uses quotation marks) and the source notes. **If any student is caught cheating on any homework assignment, the highest score the student can earn in that course is a "C".**

Important Dates

Midterm Test	Week 9 or 10
Final Exam	Refer to the notice of the Academic Affairs Office

Note: *This syllabus is tentative and may be changed or modified throughout the semester. All students will be notified and a new syllabus will be given.*

Instructor: Jingning Li **Department Head: Jingning Li**