Capital University of Economics and Business
Overseas Chinese College
Course Syllabus

Year and Semester	2023Fall
Course Name	Data Science
Course Code	MIS453
Course Type	☐ General Education (Required) ☐ General Education (Elective) ☐ Professional Course (Required) ☑ Professional Course (Elective) ☐ Basic Disciplinary Course
Course Credits	3
Course Hours	51
Prerequisites	Probability & Statistics and Manage Statistics
Instructor	Prof. Emma Zhu
Contact Information	Office: C217 Tele: (010)83951082 Email: zhuleilei@cueb.edu.cn
Office Hour	TBA
Learning Centre	TBA
Grade	2020IT
Course Time/Place	2020 IT; W: 9:55-12:20 /B208

Textbook

Robert I. Kabacoff, *R in Action, Data Analysis and Graphics with R*, Manning Publications, ISBN:.978-1935-18239-9

Reference Book

- Rachel Schutt, Cathy O'Neil. *Doing Data Science*. McGraw-Hill, O'Reilly Media, 2020.
- L. Blank. *Statistical procedures for engineering, management, and science*. McGraw Hill, New York, 1980.
- W. Feller. *An introduction to probability theory and its applications. Wiley series in probability and mathematical statistics.* Wiley, New York, third edition, 1967-1968.
- Y. G. Sinay. *Probability theory, an introductory course*. Springer-Verlag, Berlin; New York, 1992.

Course Description

Data Science is an introductory course for students majored in information system management. The main content is basic introduction to data science and algorithms, topics covered will include concept and features of data science process, jobs related with data science, basic R programming, data visualization skills, complex models includes, regression, time series, clustering, classification, principle component analysis and factor analysis as well as cases of applications in data science. Students will not only develop skills of data analysis and ability of data-driven decision making, but also scientific thinking which are all indispensable for future study and professions..

Student Learning Objectives

After completing this course, students will be able to:

Knowledge:

- describe concept and features of data science process;
- describe complex data analysis models;
- analyze data using complex models;
- apply R coding in data analysis;

Capability:

- analyze data using proper algorithms;
- develop ability of data-driven decision making by data science process;
- construct the scientific thinking and mindset, include systematic thinking, logic thinking, critical thinking and strategic thinking;

Value:

- develop the quality and morals of being objective, integrity and dedication;
- criticize the world with data-driven philosophical view.

Website Source

- https://www.coursera.org/browse/data-science
- https://www.ibm.com/topics/data-science
- https://data36.com/what-is-data-science/

Teaching Methods

This course consists of lectures, discussions, group projects, assignments, individual presentations and online activities. Students must be prepared to finish some small questions and small quiz during the class on Xuexitong application.

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, preparation of financial statements and a summary of gains from a particular course. 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%	

Evaluation criterion for presentations

Component	**Description & Requirement**
Content (50%)	Your presentation must start with a delivery of key conclusions and recommendations. It is not a recapitulation of your entire analysis. The subsequent parts of your presentation should clearly lead the audience to understand how you arrived at your conclusions and recommendations.
Coherence (10%)	You have a clearly developed message that flows naturally from your presentation. The transitions are smooth. The presentation is succinct and not choppy.
Organization (10%)	Follow the format provided in the outline. Introduce your team and the agenda you will follow. Provide handouts to the audience prior to beginning your presentation. Indicate when you would like to take questions.
Creativity (10%)	Require the use of Power Point, you can add originality to the presentation to capture and hold the audience's attention. You can also go too far in your creativity. If your presentation uses annoying or distracting sounds, for example, it negatively impacts on creativity.
Speaking skills (15%)	The criteria include: poise, clear articulation, proper volume, steady rate, good posture, eye contact, enthusiasm, and confidence. The speakers do not read (e.g., note cards, read the overhead transparencies).
Timeliness	You have 5-7 minutes to make your presentation. This is the typical amount of time that

(5%)	you can expect before a group of senior managers.

Detailed Grade Computation

	Before Midterm	After Midterm
Attendance	5%	5%
Participation	5%	5%
Homework	5%	10%
Quizzes	5%	10%
Presentation		10%
Midterm test	20%	
Final exam		20%
Total	40%	60%

Grading Policy

A+ 97-100	A 93-96	A- 90–92	B+ 87-89	B 83-86	B- 80–82
C+ 75-79	C 70-74	C- 67–69	D+ 63–66	D 62-60	F 0- 59

Exam Schedule

Midterm Test: 9th week

Final Exam: Oct. 30th – Nov. 3th

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.

Topical Course Outline

Week	Content	Platform	Homework
1	✧ Syllabus & Orientation ✧ Ch1 Introduction to data science o Data science process o Data science jobs o Data types	Classroom & XueXitong	XueXitong
2	✧ Ch2 Introduction to R o Install R and R studio o R data types o R programming	Classroom & XueXitong & R studio	
3	✧ **Summary and discussion of HW** ✧ Ch3 Data visualization o Introduction of ggplot2 o Function qplot o Function ggplot o R programming with ggplot2 ✧ **Summary and discussion of HW**	Classroom & R studio	XueXitong
4	✧ Ch4 Regression o Introduction of regression models o Evaluating model assumptions o Selecting among competing models o Case analysis o R programming with regression models ✧ **Summary and discussion of HW**	Classroom & XueXitong & R studio	XueXitong

5	✧ National Holiday		
6	✧ Ch5 Clustering o Identify cohesive subgroups o Determine the number of clusters o HAC with case analysis	Classroom & XueXi tong & R studio	
7	o PAC with case analysis o R programming with clustering analysis ✧ **Summary and discussion of HW**	Classroom & XueXi tong & R studio	XueXitong
8	✧ Ch6 Classification o Introduction of classification models o Logistic regression with case analysis o Decision trees with case analysis o Random forest with case analysis	Classroom & XueXi tong & R studio	
9	✧ Review and Midterm	Classroom & XueXi tong & R studio	XueXitong
10	o SVM with case analysis o Evaluate classification accuracy o R programming with classification models ✧ **Summary and discussion of HW**	XueXi tong & R studio	XueXitong
11	✧ Ch7 Time series o Introduction to time series o Decomposing a time series into components o Develop predictive models	Classroom & XueXi tong & R studio	
12	o Forecast future values o R programing with time series ✧ **Summary and discussion of HW**	Classroom & XueXi tong & R studio	XueXitong
13	✧ Ch8 PCA and factor analysis o Introduction to data reduction techniques o PCA with a case analysis	Classroom & XueXi tong & R studio	

14	○ EFA with a case analysis ○ R programing with PCA and EFA ✧ **Summary and discussion of HW**	Classroom & XueXi tong & R studio	XueXitong
15	✧ Presentation tutoring	Classroom	
16	✧ Presentation	Classroom	
17	✧ Final Review	Classroom & R studio	

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"**

Instructor: Leilei Zhu **Department Head: Jingning Li**