Foundation of Data Science

Department: Fudan International Summer Session

Course Code	ECON170017		
Course Title	Foundation of Data Science		
Credit	2	**Credit Hours**	36+3 tutorial hours (one credit hour is 45 minutes)
Course Nature	□Specific General Education Courses □Core Courses ☑General Education Elective Courses □Basic Courses in General Discipline □Professional Compulsory Courses □Professional Elective Courses □Others		
Course Objectives	The class, Foundations of Data Science, is designed to be a freshman level data science class that focuses on the fundamentals of data science with some primary introductions of basic machine learning algorithms near the end of the class. Instead of focusing on the theory of machine learning and data analysis, we will get started with data analysis directly. The course content is primarily based on the undergraduate course, The Foundations of Data Science, from UC-Berkeley and UC-San Diego.		
Course Description	This class will teach you how to explore data in a scientific way and show you the importance of data analysis. It will also teach you skills for programming data analysis code in Python. The topics included in the schedule adopts a breadth-first approach to give you a big picture of data science. Specifically, at the end of this course you will be able to: • Understand the basics of Python programming • Understand important statistics concepts such as sampling, hypothesis testing, and confidence intervals. • Understand experimental design to gather data • Use appropriate classification and inference tools to analyze data.		
Course Requirements: The pre-requisite of this class is basic high school algebra and an inquisitive mind. There is no requirement on prior programming experience. Each student is expected to have a computer. Either Windows or Mac is fine.			

Teaching Methods:
Lectures

Instructor's Academic Background:
Dr. Cao Yingjun received his Ph.D. in Computer Engineering from Duke University (Durham, NC, USA). His primary research interest is Computer Science Education with a focus on collaborative learning. He is also involved in more traditional research on network data analysis and distributed learning. Dr. Cao has been teaching in the Department of Computer Science & Engineering at the University of California, San Diego since 2015.
Email: yic242@eng.ucsd.edu

Course Schedule:

	Topic	**Reading**	**Lab**
1	Introduction / cause & effect	Chapters 1 - 2	Lab1: Expressions
	Expressions and data types	Chapters 3 - 4	
	Lab time		
2	Sequences	Chapter 5	Lab2: Types and sequences
	Tables	Chapter 6	
	Lab time		
3	Charts / Histograms	Chapter 7	Lab3: Arrays and Tables
	Functions and apply	Chapter 8 intro, 8.1	
	Lab time		
4	Groups / joins	8.2-8.5	lab4: Functions and visualizations
	Iterations, conditionals	9.1-9.2	
	Lab time		
5	Simulation and chance	9.3 - 9.5	Lab5: Randomization
	Sampling and empirical distributions	Chapter 10	
	Lab time		
6	Models	11.1	Lab6: Statistics and Samples
	Hypothesis testing	11.2-11.4	
	Lab time		
7	AB testing, causality	Chapter 12	Lab 7: Bootstrap
	Bootstrapping and confidence interval	Chapter 13	

	Lab time		
8	CI for hypothesis testing, center and spread	13.4, 14.1-14.2	No Lab
	Normal distribution, CLT	14.3-14.4	
	Free time		
9	Sample means, designing experiments	14.5, 14.6	Lab8: regression
	correlation and regression	15.1 - 15.2	
	Lab time		
10	Least squares, regression inference	15.3, 15.5, Chapter 16	Lab9: Regression inference
	Classification	Chapter 17	
	Lab time		
11	Decisions	Chapter 18	Lab10: Decisions
	Wrap up and finish lab 10	No reading	
12	**Final Exam**		

The design of class discussion or exercise, practice, experience and so on:
Exercise and practice

Grading & Evaluation
Your final grade will be determined via the following percentages:
Lecture participation points: 5%
Labs: 50%
Final: 45%

Important grading policies:

- You must score at least 55% on the final exam to pass the course. If you score lower than 55% on the final, you will receive an F for the course, regardless of your overall average.
- Every student should follow the policy on pair programming.
- According to Fudan University's policy, there is a threshold on the percentage of students who may receive A or A- in a class. Please keep this policy in mind.

There will be no make-up exam.

Teaching Materials & References:

- Textbook for our class will be the freely available awesome textbook, Computational and Inferential Thinking - The Foundations of Data Science", by Ani Adhikari and John DeNero.
- A reference textbook very useful for AI is Artificial Intelligence: A Modern Approach by S. Russell and P. Norvig.
- There will be a reading assignment for most of the days. It is expected that you complete the reading assignment before the start of the lecture.
- You should score at least 55% in the final exam to get a passing grade for this class, regardless of your overall percentage.
- There is an optional final project that will be treated as extra credit.