

## Foundation of Data Science

### Global Program on Economics and Finance (2021 Fall)

<b>Course Title</b>	<b>Foundation of Data Science</b>		
<b>Credit</b>	3	<b>Credit Hours</b>	48 credit hours
<b>Course Objectives</b>	<p>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.</p>		
<b>Course Description</b>	<p>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:</p> <ul style="list-style-type: none"> <li>• Understand the basics of Python programming</li> <li>• Understand important statistics concepts such as sampling, hypothesis testing, and confidence intervals.</li> <li>• Understand experimental design to gather data</li> </ul> <p>Use appropriate classification and inference tools to analyze data.</p>		
<p><b>Course Requirements:</b> 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.</p> <p><b>Prerequisites:</b> Basic algebra</p>			
<p><b>Teaching Methods:</b> Lectures and labs (part of a lecture)</p>			
<p><b>Course Schedule:</b> schedule is subject to change. The schedule assumes that each day has 4 scheduled lecture sessions (45 minutes each). The first 3 sessions are lectures and the last session is lab.</p>			
<b>Foundations of Data Science</b>			
<b>Day</b>	<b>Topic</b>	<b>Reading</b>	<b>Lab</b>
1	Introduction / cause & effect	Chapters 1 - 2	Lab1: Expressions
	Expressions and data types	Chapters 3 - 4	
	Sequences	Chapter 5	

	Lab time			
2	More about Sequences	Chapter 5	Lab2: Types and sequences	
	Tables	Chapter 6		
	Charts / Histograms	Chapter 7		
	Lab time			
3	More about Charts / Histograms	Chapter 7	Lab3: Arrays and Tables	
	Functions and apply	Chapter 8 intro, 8.1		
	Groups / joins	8.2-8.5		
	Lab time			
4	Final project proposal presentation		lab4: Functions and visualizations	
	Iterations, conditionals	9.1-9.2		
	Simulation and chance	9.3 - 9.5		
	Lab time			
5	More about Simulation and chance	No new reading	Lab5: Randomization	
	Sampling and empirical distributions	Chapter 10		
	Models	11.1		
	Lab time			
6	More about Models	No new reading	Lab6: Statistics and Samples	
	Hypothesis testing	11.2-11.4		
	AB testing, causality	Chapter 12		
	Lab time			
7	More about AB testing, causality	No new reading	Lab 7: Bootstrap	
	Bootstrapping and confidence interval	Chapter 13		
	CI for hypothesis testing, center and spread	13.4,14.1-14.2		
	Lab time			
8	Final project progress report		Lab 8: Resampling and bootstrap	
	Normal distribution, CLT	14.3-14.4		
	Sample means, designing experiments	14.5		
	Lab time			
9	More about Sample means, designing experiments	14.6	Lab 9: regression	
	correlation and regression	15.1 - 15.2		
	Least squares, regression inference	15.3, 15.5		
	Lab time			
10	More about Least squares, regression	Chapter 16	Lab10: Regression inference	

	inference		
	Classification	Chapter 17	
	Decisions	Chapter 18	
	Lab time		
11	Final review and final project Q/A session		Lab11: Decisions
	Lab time		
12	Final Project presentation		
Final Exam TBA			
<p><b>The design of class discussion or exercise, practice, experience and so on:</b></p> <p>Exercise and practice</p>			
<p><b>Grading &amp; Evaluation:</b></p> <p>Your final grade will be determined via the following percentages:</p> <p>Lecture participation points: 10%</p> <p>Labs: 45%</p> <p>Final project: 15%</p> <p>Final: 30%</p> <p><b>Important grading policies:</b></p> <ul style="list-style-type: none"> <li>• Every student should follow the policy on pair programming.</li> <li>• 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.</li> </ul> <p>There will be no make-up exam.</p>			
<p><b>Teaching Materials &amp; References:</b></p> <p>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.</p>			