

FNCE 2370: Data Science for Finance

Syllabus – Spring 2024

[Prof. Riccardo Sabbatucci](#)

**Visiting Associate Professor -- Department of Finance
The Wharton School, University of Pennsylvania**

Overview: This course will introduce students to data science for financial applications using the Python programming language and its ecosystem of packages. Students will investigate a variety of empirical questions from different areas within finance including: FinTech, asset management, corporate finance, corporate governance, personal finance, venture capital, and private equity. The course will highlight how big data and data analytics shape the way finance is practiced by focusing on problems currently confronting finance professionals.

Objective: The course objectives are threefold:

1. Improve financial decision-making with data analytics and the scientific method
2. Provide a foundation for performing data analytics in finance-related roles inside (e.g., commercial and investment banking, venture capital, private equity, asset management) and outside the financial sector (e.g., consulting, general management, corporate development, treasury).
3. Train “data translators” who can bridge the gap between technical professionals (e.g., computer scientists, statisticians) and strategic/operational professionals (finance, sales, marketing, operations, human resources, etc.)

To achieve these objectives, students will:

- *Further intellectual curiosity:* Students are encouraged to probe issues more deeply with further questions and analysis in class and on assessments.
- *Engage in collaboration:* Students are required to work in teams for some assessments.
- *Leverage diversity:* Class and student teams will reflect diversity along many dimensions including education, experience, and demographics.
- *Broaden knowledge base:* We cover a broad and diverse set of topics.
- *Reason from first principles:* Fundamental principles are emphasized, not memorization.
- *Apply the scientific method:* Questions are addressed using the scientific method.
- *Hone intuition:* We identify commonalities among seemingly disparate problems.
- *Engage with senior leaders:* The course materials are inspired by and developed in concert with practitioners, many of whom are alumni.

Classroom: We will code together in class and in small groups via “Datathons” to answer financial questions through data analysis. Most of our work will be performed in [Jupyter Notebooks](#) so please bring a Wi-Fi enabled laptop and a power adapter to every class.

Before the first class:

1. Install the [Individual Edition of Anaconda](#), Python 3.7 or higher, on your laptop. We will use [JupyterLab](#), which comes with Anaconda, but you are free to use [Google Colab](#) or an IDE with the ability to work natively with notebooks (e.g., [VS Code](#) or [PyCharm](#)).
2. For those unfamiliar with Python, please read all chapters of [A Whirlwind Tour of Python](#). You should also get acquainted with JupyterLab and Jupyter Notebooks, our development environment and web-based computing platform (i.e., where we'll be coding). See the Excel file "fnce-2370-spring-2024-python-resources.xlsx" on Canvas for links to YouTube videos and other resources.

Attendance: You can miss three classes for whatever reason without consequence. The fourth missed class lowers your grade one full level (e.g., "A" to "B"). The fifth missed class leads to being drop-failed from the class.

Assessment: Your course grade is based on three components:

1. Engagement (20%):
 - (10%) Participation:
 - In class: Ask and answer questions, comment. We count the days on which you participate at least one time. Students caught surfing the web or engaging in other non-course related activities in class will have their grade reduced one full level (e.g., "A" to "B", "B+" to "C+", etc.) for the first offense, and drop-failed from the course for the second offense.
 - Outside class: Engage on [Ed Discussion](#) (accessible through [Canvas](#)) by asking and answering questions, posting comments, sharing information, identifying errors in, or improving the quality of class materials. Your posts will be counted and assessed based on their merit.
 - (10%) Team assessments: Each teammate will grade the other teammates' effort on a scale of one to three where one corresponds to little or no effort, two corresponds to reasonable effort, and three corresponds to maximal effort.
2. Data Labs (50%):
 - For each data lab, you must submit via Canvas a fully annotated and executed Jupyter Notebook that answers the data lab questions. You may work in teams of up to but no more than *three people*. Submit executed Jupyter notebook to Canvas by 11:59 PM the day *before* discussed in class (see the schedule below)
 - Each teammate will grade the other teammates' effort on a scale of one to three where one corresponds to little or no effort, two corresponds to reasonable effort, and three corresponds to maximal effort.
3. Project (30%): same groups as for the Data Labs (max three students)
 - Phase 1: Proposal. 3–5-page writeup detailing project: Question(s), hypotheses, tests of hypotheses, and data. The project must be viable (e.g., you can access data).

- Phase 2: Preliminary results. Maximum 15-page writeup containing outline of final writeup, and preliminary results (tables and figures).
- Phase 3: Final product. One zip file containing:
 - 3-5 PDF slide deck for 5 min in-class presentation
 - Maximum 35-page writeup including all figures, tables, and references
 - Jupyter notebooks used to process data and generate results, tables, and figures. (Notebooks + data must be able to easily reproduce all results in paper so clearly document code.)
- Phase due dates are indicated in the schedule.

Pre-requisites: The formal pre-requisites for the course are FNCE 100, STAT 102, and at least one semester of programming experience, though knowledge of Python is not assumed. (Acceptable substitutes for STAT 102 include: STAT 405/422/431/432, and ESE 402.)

Resources: There is no required textbook for the course. However, the following texts are recommended and are referenced in the schedule below.

1. [*A Whirlwind Tour of Python \(Vanderplas\)*](#)
2. [*Python Data Science Handbook \(Vanderplas\)*](#)
3. [*Corporate Finance 6th Edition \(Berk and DeMarzo\)*](#)
4. [*Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools, and Techniques to Build Intelligent Systems 3rd Edition \(Geron\)*](#)

The first two books are freely available as web books. The third book is the text for many sections of FNCE 100 and is included only for those wanting a refresher on different finance concepts. The last book is relatively inexpensive and a nice non-technical introduction to machine learning with lots of sample code. I have also compiled a small database of web sites and YouTube videos – all freely accessible – in an Excel file, “fnce-2370-spring-2024-python-resources.xlsx.”

Rules: All the materials (slides, data labs, Jupyter notebooks, video recordings, data, etc.) are under copyright. No part of any material may be reproduced, digitized, photocopied, stored in a retrieval system, used in a spreadsheet, or otherwise reproduced, posted, or transmitted in any form or by any means without Michael R. Roberts’ and Riccardo Sabbatucci’s permission. Your enrolment in the course signifies your agreement with these rules and to the [University of Pennsylvania Code of Student Conduct](#).

Meeting Times: It turns out that Zoom office hours are much more effective because we can easily share screens. Of course, I am happy to meet in person when necessary. Just contact me via Ed Discussions on Canvas.

Instructor	Class times	Class location	OHS
Riccardo Sabbatucci	T/T: 08:30 - 10:00	VANC B10	Tu: 15:00 - 16:00 (Zoom)
	T/T: 10:15 - 11:45	VANC B10	Tu: 16:00 - 17:00 (Zoom)

Students must attend their assigned section but may attend any office hours

Office hours will be virtual to ease screen sharing and respect social distancing.

TAs	Recitation times	Recitation location	OHS
Claudia Pinilla	F: 10:15-11:15 and 11:30-12:30	SHDH 213	

Recitations are not required, and students may attend either one.

Office hours will be virtual to ease screen sharing and respect social distancing.

Tentative Schedule

Class #	Date	Day	Topic	Classroom materials	Assessment due	Readings
1	18-Jan	Thu	Course intro / Scientific Method / Data Science Workflow			<i>WTP</i> : <i>PDSH</i> : Ch.2
2	23-Jan	Tue	Risk and Return	Nb: 12-stocks-long-run-prep	Form lab/project teams	<i>PDSH</i> : Ch.4; <i>CF</i> : 9.1, 10.1 - 10.4
3	25-Jan	Thu	Lab: Stocks for the Long Run	Lab workup and notes	Read lab writeup	
4	30-Jan	Tue	Microfinance	Nb: 10-microfinance		<i>CF</i> : 10.5 - 10.8, 11.7 - 11.8, 12, 13.1
5	1-Feb	Thu	Microfinance	Nb: 10-microfinance		Statsmodels user guide: linear regression models
6	6-Feb	Tue	Factor models	Nb: 11-factor-models		<i>CF</i> : 13.6, 13.7, 13.A
7	8-Feb	Thu	Lab: Riffled AI	Lab workup and notes	Lab Notebook	
8	13-Feb	Tue	Asset allocation	Nb: 10-crypto-investing-prep	Project: Proposal	<i>CF</i> : 11.1 - 11.5
9	15-Feb	Thu	Asset allocation	Nb: 10-crypto-investing-prep		
10	20-Feb	Tue	Asset allocation	Nb: 10-crypto-investing-prep		
11	22-Feb	Thu	Lab: Crypto Investing	Lab workup and notes	Lab Notebook	
12	27-Feb	Tue	Inflation and corporate profits	Nb: 15-inflation-profit		<i>PDSH</i> : Ch.3; <i>CF</i> : 2.1 - 2.4, 2.6, 8.1, 8.2
13	29-Feb	Thu	Inflation and corporate profits	Nb: 15-inflation-profit		
SPRING BREAK						
14	12-Mar	Tue	Inflation and corporate profits	Nb: 15-inflation-profit		<i>CF</i> : 28.1 - 28.4, 28.6
15	14-Mar	Thu	Lab: Indigo Acquisition	Lab workup and notes	Lab Notebook	
16	19-Mar	Tue	Airbnb - hedonic pricing	Nb: 10-airbnb-wrangling-split	Project: Preliminary Results	<i>ML</i> : Ch.1,2,4
17	21-Mar	Thu	Airbnb - hedonic pricing	Nb: 11-airbnb-eda		
18	26-Mar	Tue	Airbnb - hedonic pricing	Nb: 12-airbnb-ml		
19	28-Mar	Thu	Lab: Diamonds	Lab workup and notes	Lab Notebook	
20	2-Apr	Tue	Target marketing	Nb: 10-bank-marketing		<i>ML</i> : Ch.3
21	4-Apr	Thu	Target marketing	Nb: 10-bank-marketing		
22	9-Apr	Tue	Target marketing	Nb: 10-bank-marketing		
23	11-Apr	Thu	Lab: Raterx	Lab workup and notes	Lab Notebook	
24	16-Apr	Tue	Guest Speaker			
25	18-Apr	Thu	Project presentations	Project powerpoint	Project: Final Submission	
26	23-Apr	Tue	Project presentations	Project powerpoint		
27	25-Apr	Thu	Project presentations	Project powerpoint		
28	30-Apr	Tue	Project presentations	Project powerpoint		

Abbreviations:

[WTP: Whirlwind Tour of Python \(Vanderplas\)](#)

[PDSH: Python for Data Science Handbook \(Vanderplas\)](#)

[CF: Corporate Finance \(Berk and DeMarzo\), 6th edition](#)

[ML: Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow \(Geron\), 3rd edition](#)