FNCE 237: Data Science for Finance *Tentative* Syllabus – Fall 2022 <u>Prof. Michael Roberts</u> 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 <u>Jupyter Notebooks</u> so please bring a Wi-Fi enabled laptop and a power adapter to every class.

Before the first class:

- 1. Install the <u>Individual Edition of Anaconda</u>, Python 3.7 or higher, on your laptop. We will use <u>JupyterLab</u>, which comes with Anaconda, but you are free to use <u>Google Colab</u> or an IDE with the ability to work natively with notebooks (e.g., <u>VS Code</u> or <u>PyCharm</u>).
- 2. For those unfamiliar with Python, please read Chapters 3-9 and 12 (they're very short and basic) of <u>A Whirlwind Tour of Python</u>. You should also get acquainted with JuptyerLab and Jupyter Notebooks, our development environment and web-based computing platform (i.e., where we'll be coding). See the Excel file "fnce-237-fall-2022-python-resources.xlsx" on Canvas for links to YouTube videos and other resources.

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

Assessment: Your course grade is based on three components:

- 1. Engagement (20%):
 - o (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 <u>Ed Discussion</u> (accessible through <u>Canvas</u>) 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.
 - 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%):
 - 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. For each phase,

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. <u>A Whirlwind Tour of Python (Vanderplas)</u>
- 2. <u>Python Data Science Handbook (Vanderplas)</u>
- 3. *Corporate Finance* 5th *Edition* (*Berk and DeMarzo*)
- 4. <u>Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow: Concepts, Tools,</u> <u>and Techniques to Build Intelligent Systems 2nd Edition (Geron)</u>

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-237-fall-2021-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' permission. Your enrolment in the course signifies your agreement with these rules and to the <u>University of Pennsylvania Code of Student Conduct</u>.

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

Instructor	Class times	Class location	OHs
Michael Roberts	M/W: 10:15 - 11:45	JMHH F60	<u>M: 3:15 - 4:15 PM (Zoom)</u>
	M/W: 1:45 - 3:15	JMHH F60	<u>Th: 5:00 - 6:00 (Zoom)</u>

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
Daniel Mckeon	F: 10:15 - 11:15	SHDH 215	T: 4:00-5:00 (Zoom)
Gabriel Tettamanti	F: 1:45 - 2:45	SHDH 215	W: 5:15-6:15 (Zoom)

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	D		T = 1 = 1		• • • • • • • • • • • • •	D ecentions
	0/21	Day		Classroom Materials	Assessment due	Reading
1	0/31	vv	Scientific Method			PDSH. 2
			Data Science Workflow			
	0/5	14	Labor day	No class		
2	9/5		Diale & roturn	No class	Form lob/project to amo	
2	9/1	vv	RISK & IELUITI	ND-12-Stocks-long-lun-piep	Form lab/project learns	PDSH.4
2	0/12	14	Lab: Stocks for the Long Bun	Lab workup and notae	Pood Job writeup	61. 3.1, 10.1 - 10.4
3	9/12	101	Microfinance	Nh: 10-microfinance	Read lab writeup	
4	5/14	vv	Micronnance	ND. 10-micromatice		13.1,
5	9/19	М	Microfinance	Nb: 10-microfinance		Statsmodels users guide: Linear Regression Models - Ordinary Least Squares
6	9/21	W	Factor models	Nb: 11-factor-models		CF: 13.6, 13.7, 13.A
7	9/26	М	Lab: Riffled Al	Lab workup and notes	Lab Notebook	
8	9/28	W	Asset allocation	Nb: 10-crypto-investing-prep	Project: Proposal	CF: 11.1 - 11.5
9	10/3	М	Asset allocation	Nb: 10-crypto-investing-prep		
10	10/5	W	Asset allocation	Nb: 10-crypto-investing-prep		
11	10/10	М	Lab: Crypto Investing	Lab workup and notes	Lab Notebook	
12	10/12	W	Inflation and corporate profits	Nb: 15-inflation-profit		PDSH: 3
						CF: 2.1 - 2.4, 2.6, 8.1, 8.2
13	10/17	М	Inflation and corporate profits	Nb: 15-inflation-profit		
14	10/19	W	Inflation and corporate profits	Nb: 15-inflation-profit		CF: 28.1 - 28.4, 28.6
15	10/24	М	Lab: Indigo Acquisition	Lab workup and notes	Lab Notebook	
16	10/26	w	Airbnb - hedonic pricing	Nb: 10-airbnb-wrangling-split	Project: Preliminary results	ML: 1, 2, 4
17	10/31	М	Airbnb - hedonic pricing	Nb: 11-airbnb-eda		
18	11/2	W	Airbnb - hedonic pricing	Nb: 12-airbnb-ml		
19	11/7	М	Lab: Diamonds	Lab workup and notes	Lab Notebook	
20	11/9	W	Target marketing	Nb: 10-bank-marketing		ML: 3
21	11/14	М	Target marketing	Nb: 10-bank-marketing		
22	11/16	W	Target marketing	Nb: 10-bank-marketing		
23	11/21	М	Lab: Raterx	Lab workup and notes	Lab Notebook	
	11/23	W	Thanksgiving	No class		
24	11/28	М	Guest speaker			
25	11/30	W	Project presentations	Project powerpoint	Project: Final submission	
26	12/5	М	Project presentations	Project powerpoint		
27	12/7	W	Project presentations	Project powerpoint		
28	12/12	W	Project presentations	Project powerpoint		

Reading Abbreviations:

WTP: Whilrwind Tour of Python (Vanderplas)

PDSH: Python for Data Science Handbook (Vanderplas)

CF: Corporate Finance (Berk and DeMarzo) ML: Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow (Geron)

Abbreviations:

WTP: Whirlwind Tour of Python (Vanderplas)

PDSH: Python for Data Science Handbook (Vanderplas)

CF: Corporate Finance (Berk and DeMarzo)

ML: Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow (Geron)