FNCE 237: Data Science for Finance

Tentative Syllabus

Prof. Michael Roberts

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 (e.g., Dask, Matplotlib, Numpy, Numba, Pandas, SciPy, Scikit-Learn, StatsModels). Students will investigate a variety of empirical questions from different areas within finance including: FinTech, asset management, corporate finance, corporate governance, 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 objective is twofold:

1. Illustrate how data analytics can improve financial decision-making, and
2. Provide students with a foundation for performing data analytics in finance-related roles both inside the financial sector (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).

To achieve these objectives, the class will help students:

- *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 and encouraged to do so for others.
- *Leverage diversity*: Student teams will reflect diversity along a number of dimensions including education, experience, and demographic.
- *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*: Through exploration of a number of diverse projects students will develop intuition for working with and analyzing data.
- *Engage with senior leaders*: The course materials are inspired by and developed in concert with financial executives, several of whom have agreed to deliver guest lectures.

Classroom: We will often work together as a 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.

Assessment: Your course grade is based on four components:

1. Five quizzes account for 40% of the course grade. You are encouraged, but not required, to collaborate with your colleagues. Your lowest quiz score will be dropped.
2. Five data labs (see below) account for 40% of the course grade. You will work in teams of three on the labs. Your lowest lab score will be dropped.
3. Class participation, as measured by quality of course engagement, comprises the final 20% of the course grade.

The structure of each data lab and the final project will be a typical data science workflow executed in a Jupyter notebook:

1. Clear statement of the financial question and any background motivation or research.
2. Data acquisition
3. Cleaning and pre-processing the data
4. Exploratory data analysis (EDA)
5. Modeling
6. Inferences and conclusions

Much like cases, data labs contain subjective elements. They will be assessed based on their completeness, accuracy, readability, performance, and consistency with financial principles.

Pre-requisites: The formal pre-requisites for the course are FNCE 100 and STAT 102. Some programming experience is helpful though knowledge of Python is *not* assumed.

Preliminaries: We will be relying on the Conda distribution of Python version 3.5 or higher. Conda comes in two flavors: Miniconda and Anaconda. You should use the Anaconda distribution, which includes most of the packages that we will be using. Please come to the first class with the Anaconda distribution of Python installed on your laptop.

A nice resource for setting up your Python environment with Anaconda and deploying a Jupyter Notebook can be found at Quantitative Economics. For this course, you can ignore the discussion of integrated development environments (IDEs) and text editors.

Resources: There is no required textbook for the course. I recommend Berk and DeMarzo's, *Corporate Finance*, for finance reference. For programming, I recommend the following books by Jake Vanderplas, both of which are freely available:

1. *A Whirlwind Tour of Python*
2. *Python Data Science Handbook*

Additional resources and references are included with every data lab for students wishing to explore specific topics more deeply. All other material will be posted to Canvas.

Data labs: The class will select data labs from the list below – subject to breadth requirements and time constraints.

#	Lab description	Data	Finance topics
1	The Cryptocurrency landscape	Kaiko Cryptocurrency database	FinTech, cryptocurrency, risk exposure, investment opportunities
2	Algorithmic trading	CRSP: *msf*	Stock markets, market microstructure
3	High frequency trading	TAQ	Market microstructure, transaction costs, (Really) big data
4	Investment exit: The path to IPO	S&P Compustat: SEC S-1 filings	Capital markets, IPOs, firm performance, market expectations, venture capital, PE growth equity
5	Corporate loans and the board of directors	SEC filings: Material contract attachments	Corporate governance, bank lending, covenants
6	Corporate valuation	S&P Compustat	Corporate valuation, free cash flow, cost of capital, DCF
7	The syndicated loan market	Thompson-Reuters: *Dealscan*	Debt financing, loan structuring, commercial banking, capital markets
8	Optimal capital structure	S&P Compustat: *funda*	Capital structure, Modigliani & Miller, market imperfections
9	Predicting credit ratings	S&P Compustat: *funda, adsprate*	Credit risk, liquidity, solvency, credit ratings
10	Stock performance over the long-run	CRSP: *msix*	Personal finance, time-diversification, investment management
11	Estimating the equity cost of capital	CRSP: *mse, msf* Fama-French: *Research data factors*	Equity cost of capital, CAPM
12	Stock betas and volatility	TAQ CRSP: *dsf, msf*	High frequency data, beta estimation
13	Industrial composition of US economy in 20th century	Moody's S&P Compustat	Industrial composition and financial markets, financial characteristics of firms
14	Financial engineering	OptionMetrics	Derivatives, call & put options, Black-Scholes
15	Predicting the future	OptionMetrics	Derivatives, state prices, capital budgeting
16	What do CEOs get paid and why?	S&P Execucomp	Financial incentives, executive compensation, corporate governance
17	Socially (ir)responsible investing	CRSP	Asset management, portfolio formation, risk exposure
18	Experimental design from the top		Guest Instructor

Rules: This class will utilize proprietary data. You may not post, distribute, or share any data used in this course. Your enrolment signifies your agreement to this rule and to the University of Pennsylvania Code of Student Conduct.

Any regrade request on an assessment must be made within one week of receiving the grade for that assessment. Requests made outside that window will not be considered.

You may miss up to two classes for any reason with the exception of guest lectures for which attendance is mandatory. A third absence will lower your course grade one full level (e.g., “A” to “B”, “B+” to “C+”, etc.). A fourth absence will result in being drop-failed from the course.

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.