**Fall 2025 FRE-GY-7831 Financial Analytics & Big Data**

**Song Tang, st290@nyu.edu, 646-283-4578**

Overview:

This course provides students with a conceptual and technical understanding of the role of relational and big data databases in financial analytics. It explores how data engineering, management, and analysis enable modern financial decision-making. Through a combination of lectures, hands-on exercises, and projects, students will learn to integrate computer-aided data management, machine learning, and large-scale data processing techniques to solve practical problems in data-driven financial environments.

Schedule of Classes:

| Week | Topics |
|---|---|
| 1 | Market Data & JSON |
| 2 | Relational Database & SQLite3 |
| 3 | Lab 1: JSON-SQLite3 Lab |
| 4 | Big Data and Hadoop + Quiz 1 |
| 5 | Effective Python for Finance |
| 6 | Lab 2: Amazon AWS Cloud Hortonworks Hadoop |
| 7 | ML Market Prediction + Quiz 2 |
| 8 | Final Exam |

Assessment:

The assessment will be done as follows:

- Homework Assignments 30%
- Quizzes 20%
- JSON-Sqlite3 & Big Data Labs, 20%
- Final Exam, 30%

Development Environment:

(1) C++ Integrated Development Environment:
   - Microsoft Visual Studio + Amazon AWS EC2 Linux Virtual Machine

(2) Python Integrated Development Environment:
   - Visual Studio Code or Anaconda Spyder

(3) Market Data Feed:
   - EOD Financial Data API, $29 for 2 months (after a 50% student discount)

Textbooks:

Market Data Explained: A Practical Guide to Global Capital Markets Information. (The Elsevier and Mondo Visione World Capital Markets), Mark Alvarez, Butterworth-Heinemann, 2006, ISBN-10: 0750680555

Robert Nisbet, et. al. (2009) Handbook of Statistical Analysis & Data Mining Applications. Elsevier/Academic Press. San Diego, California. ISBN: 978-0-12-374765-5.

Data Science for Business: What you need to know about data mining and data-analytic thinking. Foster Provost and Tom Fawcett, O'Reilly Media, 2013, ISBN-10: 1449361323

Steve LaValle, et. al. (2011) Big Data, Analytics and the Path from Insights to Value. MIT Sloan Management Review. Winter 2011, Vol. 52, No. 2.

Boris Kovalerchuk, Evgenii Vityaev. (2000) Data Mining in Finance: Advances in Relational and Hybrid Methods. Springer, ISBN: 0792378040.

Course Topics:

I. **Market Data:** Due to technical advancement and industry innovations such as high-frequency trading, the volume of financial data will continue growing exponentially. Topics covered in this unit:
    - What is Financial Data?
        - Equities: stock quotes, daily open, high, low, close prices, trading volume, VWAP, and IPOs.
        - Companies: key company fundamentals, financial statements and ratios, analyst stock recommendations and ratings, analyst earnings estimates, and financial, economic, and business news.
        - Mutual Funds & ETFs: historical and daily end-of-day closing NAVs, ETF, open-ended mutual funds, and money market fund data.
        - Fixed Income & Credit: bond price and yield data, US Treasury and LIBOR-based swap and forward rates, and global interbank interest rates and the official BBA LIBOR.
        - Futures & Options: daily end-of-day prices, quotes, volume, and open interest for all options and futures.
        - Forex & Metals: quotes for currency exchange rates, precious metal spot prices, and global currency pairs.
        - Indices & Markets: index values for more than 10,000 U.S. and international indices, and global economic calendars of scheduled Treasury and economic events and announcements.
    - How is Financial Data Delivered?
        - Delivery frequency: real-time, delayed, conflated, or end-of-day (EOD).
        - Delivery method: streaming format, snapshot files, or EOD files.

- Delivery transportation: broadcast, multicast, satellite, private line, VPN, or Internet.
- Delivery format: encoded heavily to optimize performance, or in simple formats to simplify databasing.
- Normalization: a vendor collects from sources all around the world and then translates all those formats into a single format.
- Reliability: The high availability of data is a primary concern in the financial markets.
- Value Added Services: data value can be improved by adding related services such as listing information, sharing data, fundamental data, etc.

II. **Relational Database**: This unit will introduce students to the concepts of relational database management systems and relational database models. Students will learn how to use relational databases in trading via pair trading implementation.

Topics covered in this unit:

- Relational Database
  - DBMS
  - Relational Database Model
  - Relational Algebra
  - Normalization
  - E-R Relationship
  - Structured Query Language
- Pair Trading Implementation
  - Restful
  - Sqlite3 Integration
  - Pair Trading System
    - Condition and Assumption
    - Pair Trading Algorithm
    - Database Implementation Details

III. **JSON-Sqlite3** Using libcurl to pull daily and intraday trading data from Unicorn Market Data services in JSON format, parse the market data, create and populate corresponding data structures, and create multiple tables to persist the market data in a SQLite3 relational database.

Topics covered in this unit:

- What is JSON?
- JSON Object Syntax
- JSON Examples
- JSON Arrays
- JSON Data Types
- JSON Usage
- Market Data in JSON Format

- C++ Program, FRE7831_MarketData_JSON
- Sqlite3
- Add Sqlite3 into FRE7831_MarketData_JSON
- Create a Sqlite3 relational database.
- Create multiple tables with Primary Key & Foreign Key constraints
- Populate the market data from STL data structures into tables for persistence.

IV. **Big Data in Finance:** While industry-structured data is growing in size and scope, it is the world of unstructured data that is emerging as an even larger and more important data source. IBM's Big Data Work survey indicates that most financial organizations are currently in the early stages of big data planning and development efforts.
Topics covered in this unit:
- What is Big Data?
- Big Data Challenges in Financial Markets
- Structured vs. Unstructured Data
- Main Big Data Technologies & Hadoop
- Harvest Financial Information Using Big Data
  - Hortonworks Sandbox and HDP
  - Aggregating 10 Years of Raw Stock Ticker Data from NYSE
  - Enriching the Data Model with Unstructured Data from the Internet
  - Interactive Visualization

V. **Effective Python for Financial Data Analysis:** This section will introduce Python with a focus on Python packages for effective numerical and data analysis.
Topics covered in this unit:
- Default data structures and improved versions, defaultdict, namedtuple
- 3rd party packages for data analysis
- NumPy – Numerical Python
- SciPy - tools and functions for scientific computing
- Visual Finance via Matplotlib and seaborn
- Pandas and effective data management
- Manipulate relational database.
- Using SQLite in Python

VI. **Amazon AWS Cloud & Cloudera Hadoop:** This section is a hands-on lab for students to practice Big Data databases in the Amazon AWS Cloud.
Topics covered in this unit:
- Create an AWS Account and add an IAM User
- Sign in as IAM User
- Create EC2 t2.xlarge instance
- Download Cloudera Sandbox from GitHub

- Install Docker on your EC2 instance
- Install HDP 2.6.5 through docker
- Access Ambari to make sure all Cloudera processes are running.
- Copy the stocks.tar file from your local computer to Hortonworks
- Run DDL and DML in Hive View to create and populate tables
- Download and install the ODBC driver on your local computer
- Set up ODBC driver to connect your Hortonworks env
- Run a Python program on your local computer to use data in a Cloudera env

VII. **ML Market Prediction:** We will create a Python application to predict future stock price movement through an ML application on technical analysis.
Topics covered in this unit:
- Google deep learning framework TensorFlow
- Market data integration
- Data analysis visualization
- Python Technical Analysis library
- ML Application for Stock Price Movement Prediction.

Letter Grades:
Letter grades for the entire course will be assigned as follows:

| Letter Grade | Points | Percent |
|---|---|---|
| **A** | 4.00 | 93.33% |
| **A-** | 3.67 | 90.00% |
| **B+** | 3.33 | 86.67% |
| **B** | 3.00 | 83.33% |
| **B-** | 2.67 | 80.00% |
| **C+** | 2.33 | 76.67% |
| **C** | 2.00 | 70.00% |
| **F** | 0.00 | 0.00% |

For academic integrity, there will be no grading curving, additional assignments, makeup exams, or grade adjustments once all scheduled classwork is completed.

In addition, this course assumes that work submitted by students will be generated by the students themselves, working individually or in groups as directed by class assignment instructions. Any classwork generated by anyone other than the students (by other students, by a company, or by using generative AI tools in ways that violate this course policy) will be considered a breach of the university's Academic Integrity policy.

Policies:

**Academic Misconduct**

A. Introduction: The School of Engineering encourages academic excellence in an environment that promotes honesty, integrity, and fairness, and students at the School of Engineering are expected to exhibit those qualities in their academic work. It is through the process of submitting their own work and receiving honest feedback on that work that students may progress academically. Any act of academic dishonesty is seen as an attack upon the school and will not be tolerated. Furthermore, those who breach the school's rules on academic integrity will be sanctioned under this Policy. Students are responsible for familiarizing themselves with the School's Policy on Academic Misconduct.
B. Definition: Academic dishonesty may include misrepresentation, deception, dishonesty, or any act of falsification committed by a student to influence a grade or other academic evaluation. Academic dishonesty also includes intentionally damaging the academic work of others or assisting other students in acts of dishonesty. Common examples of academically dishonest behavior include, but are not limited to, the following:
    1. Cheating: intentionally using or attempting to use unauthorized notes, books, electronic media, or electronic communications in an exam; talking with fellow students or looking at another person's work during an exam; submitting work prepared in advance for an in-class examination; having someone take an exam for you or taking an exam for someone else; violating other rules governing the administration of examinations.
    2. Fabrication: including but not limited to, falsifying experimental data and/or citations.
    3. Plagiarism: Intentionally or knowingly representing the words or ideas of another as one's own in any academic exercise; failure to attribute direct quotations, paraphrases, or borrowed facts or information.
    4. Unauthorized collaboration: working together on work that was meant to be done individually.
    5. Duplicating work: presenting for grading the same work for more than one project or in more than one class, unless express and prior permission have been received from the course instructor(s) or research adviser involved.

6. Forgery: altering any academic document, including, but not limited to, academic records, admissions materials, or medical excuses.

**Disability Disclosure Statement**

Academic accommodation is available for students with disabilities. Please contact the **Moses Center for Students with Disabilities** (212-998-4980 or mosescsd@nyu.edu) for further information. Students who are requesting academic accommodations are advised to reach out to the Moses Center as early as possible in the semester for assistance.

**Inclusion Statement**

The NYU Tandon School values an inclusive and equitable environment for all our students. I hope to foster a sense of community in this class and consider it a place where individuals of all backgrounds, beliefs, ethnicities, national origins, gender identities, sexual orientations, religious and political affiliations, and abilities will be treated with respect. It is my intent that all students' learning needs be addressed both in and out of class and that the diversity that students bring to this class be viewed as a resource, strength, and benefit. Please feel free to speak with me if this standard is not being upheld.