Spring 2019 FRE-GY-7831 Financial Analytics & Big Data
Song Tang, st290@nyu.edu, 646-283-4578

Overview:

The purpose of this course is to provide students a greater appreciation, both conceptual and technical of the uses of relational databases and big data databases and their engineering of financial analytic. The course covers a number of themes and practical problems that combine computer aided data management, data analysis, financial issues and their applications in a data based financial environment.

Schedule of Classes:

Week	Topics
1	Market Data & Bloomberg API
2	Relational Database Term Project: Pair Trading (Part 1)
3	BLPAPI-RDB Lab
4	Big Data and Hadoop Term Project: Pair Trading (Part 2)
5	Big Data Lab: Microsoft Azure Cloud & Hortonworks Hadoop
6	Python for Finance Term Project: Pair Trading (Part 3)
7	Final Exam Term Project: Pair Trading (Due)

Assessment:

Students will have homework assignments and final exams. The assessment will be done as the following. The homework assignments will require significant amount of coding.

- Quizzes 20%
- Homework Assignments, 20%
- Big Data Lab, 10%
- Final Exam, 25%
- Team Project, 25%

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:
   o What are 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 currencies 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.
   o How are 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 of those formats into a single format.
- Reliability: high availability of data is a primary concern in the financial markets.
- Value Added Services: data value can be improved by adding on related services such as listing information, share data, fundamental data, etc.

II. **Relational Database & Pair Trading**: This unit will introduce students the concepts of relational database management systems and relational database model. Students will learn how to use relational database in trading via the 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. **BLPAPI-RDB** Market Data Application Programming Interfaces (API): the market data from data vendors provide data-neutral, thread-aware access for requesting and providing market information as well as some generic configuration and logging capabilities. For example, Bloomberg Open Market Data Initiative, and Reuters Market Data System (RMDS) and Open Message Model (OMM) which offers an ultra-low latency and high throughput data delivery option. Topics covered in this unit:
  - The Bloomberg API (BLPAPI)
    - Typical Application Structure
    - Application Example for retrieving historical market data and real-time data
  - Database design for market data
    - Microsoft JET Engine
    - Table Design and ER relationship
    - Populate market data via BLPAPI

IV. **Big Data in Finance:** While the 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 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 Internet
  - Interactive Visualization

V. **Microsoft Azure Cloud & Hortonworks Hadoop:** This section is a hands-on lab for students to practice Big Data database in Microsoft Azure Cloud.
Topics covered in this unit:
- Start virtual machines in Microsoft Azure Cloud
- Launch HDP and start Hue
- Create Big Data database
- Connect to your Python program to your Big Data database in Azure Cloud via ODBC
- Code your Pythonn program to analyze data retrieved from your Big Data database

VI. **Python for Financial Data Analysis:** This section will introduce Python with focus on Python packages for numerical and data analysis.
Topics covered in this unit:
- Data structure for collections
- 3rd party packages for data analysis
- NumPy – Numerical Python
- SciPy - tools and functions for scientific computing
- Visual Finance via Matplotlib
- Pandas - powerful data analysis toolkit
- Manipulate relational database
- Using SQLite in Python