NYC Data Science Capstone Project

How do firms hedge their production risks? This has been an important question for both the Wall Street and academia. Our task is to use state-of-art data science techniques to collect the hedging positions of Oil/Gas Companies and build a quantitative model of the industry's risk management practices. This is an opportunity for young data scientists to solve the real world problem and develop your expertise in the interdiscipline between big data and finance.

Data collection (Progress: 80% done):

- Scrape the following data from Oil/Gas companies' SEC filings.
    - Tables regarding hedging positions (instrument, maturity, volume, and price)
    - Tables regarding productions (volume)
- Collect oil and NG futures prices
- Construct the key financial accounting variables for all oil/gas companies)
- Extract credit rating for all oil/gas companies
- Assemble data on key macroeconomics indicators

Processing scraped data (Progress: 20% done):

- Transform tabular data from each company (inconsistent format) into a large panel that include the hedge positions of oil/gas companies between 2006 and 2018
- Identify the hedge-to-production ratio and the average hedge price for each company-year or company-quarter.

Data exploratory analysis (Progress: 0% done):

- Potential features including firm-level financial variables, macroeconomic variables, commodity prices (e.g., the industry total oil/NG production, the level of commodity price, the term structure of commodity prices, the industry debt level, interest rate, inflation rate, consumer confidence index, automobile sales)
- At the industry and firm levels, how does hedge ratio correlate with the features above

Modeling (Progress: 0% done):

- Build a quantitative model to investigate what features explain the movements in the hedge-to-production ratio at the industry and the firm levels.
- Build a quantitative model to investigate to which the extent stock prices of oil/gas companies can be explained by the commodity prices and the hedge-to-production ratio. Cross validation of the model is expected

- Out-of-sample tests of your model: Construct a novel trading strategy based on the newly constructed hedging ratios and examine whether this new strategy generates superior returns.