GA 3, CIS 3100: Using Big Data in Financial Analysis

Overview of Our Class Project
Big topics we will discuss, through an applied project involving statistical data analysis, financial statements, relational databases, and CFA-grade analysis techniques. In the project the class will take on the role of an investment firm, recommending stock investments after pursuing a rigorous statistical inspection of companies which compete with each other.

This term in 3100, we are aiming to combine the work of all class members in analyzing all 11 sectors of the US Economy, as laid out in these 11 indices, by CNBC:

- US Markets, Sector Watch, CNBC corp., https://www.cnbc.com/sectors/

Students will form teams around an industry sector of their choosing, then identify approximately 30 companies which occupy a galaxy, or common area of economic output. The companies your team should study, over the remaining course of the term should do the same things, compete against each other, even indirectly. Hence, a suitable galaxy could be 'video game companies,' or 'automobile manufacturers.' Another way to define a galaxy of companies is to declare them to be in the 'petroleum industry,' or similarly broad area, then collect firms of different sizes which generally explore, drill and/or distribute oil products.

The objective for your team's work is to identify high performing companies in your chosen economic sector. By the close of the term the class will attempt to identify a basket of the highest performing corporations in the broader economy, filing 10Q statements with the US Securities and Exchange Commission. Where possible, if you wish to include firms which are regulated by the US SEC (for example, a car maker in China, or an oil business in India), there are ways to incorporate them in their analysis, as we will discuss, below.

Using a Methodology to Analyze your Galaxy

Gathering Company Financial Statements
Each team will collect 10Q reports filed by each company, with the US SEC EDGAR system. You can acquaint yourself with this system by visiting its company search page at

- https://www.sec.gov/edgar/searchedgar/companysearch.html

Here, analysts will acquire recent 10Q filings, in spreadsheet form. Each analyst should cover approximately 5 different firms, transforming the spreadsheet so that it may be imported into an Access Database. This and other tutorials will be posted to assist you:

- https://www.youtube.com/watch?v=X28A_jnlm0k&feature=youtu.be

The database should assist your team in collecting the content of many financial statements. The year, quarter or other timestamp should also be present as you import data. This process will require each team to design tables which mirror the columns of the financial statement, with

other primary key attributes. Once large accumulations of statements have taken place (one per firm, per quarter or year), Excel is to be used to derive common ratios in financial analysis:

- Activity
- Liquidity
- Solvency
- Profitability
- Valuation
- Credit

As a class we will see how many of these commonly accepted analyses we can do, over the course of the term, with the intent of identifying firms which embody credit worthiness, operating efficiency, and value for shareholders. In order to locate the column you wish to utilize in your ratio, you may experiment with the Index Match feature in excel:

- https://www.youtube.com/watch?v=ontXHp9cwOQ&feature=youtu.be

Ways and Means

Access and Excel should provide your team with the data organization and statistical capabilities. In addition to this, a suite of tools should evolve which serve other collaboration goals, not limited to:

- Github, for team collaboration
- Google Drive, for document synchronization
- Office 365, offering real time collaboration on spreadsheets and email
- Messaging apps, which help integrate team members over mobile devices