Taniya Adhikari

DSC 540

Project Milestone 1

**US Stocks – Company and Industry Analysis**

The 3 datasets that I will be using are below:

* **CSV File** Downloaded from[US Company Information](https://www.kaggle.com/marketahead/all-us-stocks-tickers-company-info-logos?select=companies.csv)

This CSV file contains a list of Ticker Symbol, with company name, industry name, description, stock exchange name and market Cap and other informational data. The list was generated in 2018 so it needs updated market cap for the company.

* **Website –** [TradingView Screener](https://www.tradingview.com/screener/)

This website has a screener that provides data in a tabular format and can be filtered based on criteria. It contains Ticker symbols, technical indicators, and fundamental indicators with other information.

* **API –** [Financial Modeling Prep](https://financialmodelingprep.com/developer/docs)

This API gives out daily historical prices, change and volume and historical DCF pricing value for all Ticker symbols, and this API also gives historical growth of the company as well as popular news for Ticker symbol

All three data are related by Ticker symbol of the company. Ticker symbol in CSV file has one to many relationships to Website and API data by Ticker symbol.

For this Project, I will have to format the CSV file for any duplicates or missing values using python pandas library. I will have to rename few columns as well, because some columns also overlap from other data sources. It has mainly, information about the company and contains mostly string data. I will need to do lot of data cleaning in this. For website data, I will be selecting columns for fundamental indicators (P/E, ROE, EPS etc.,), company’s performance, overview information and some technical indicators. All indicators are numerical data types however some indicators are used for trend analysis, so it will need Boolean conversion. I will also need to filter the data for Market CAP above 2B. I have two reasons for this: 1.) Since stock data changes real-time, market capitalization can also change overtime, in some cases it changes weekly but companies over 1B are highly unlikely to change in weekly or monthly. 2.) Dataset for all ticker symbols can be too long 8000+ and most of the small cap companies are highly volatile to calculate intrinsic value of the stock, all market analysis is performed to avoid or minimize risk, therefore I am only interested in creating dataset for companies with low volatility. Some of the libraries I may need is Requests, html\_table\_parser, pprint, BeautifulSoup to perform web scraping. From API I intend to get daily historical prices, change, volume and DCF value for each Ticker Symbol up to 1 year and historical growth of the company as well as some recent stock news data for sentiment indicator. This project is a small attempt to build trading strategies using, historical prices, financial indicators, and stock news data. Since stock prices change real-time, I will only include historical data for prices and news.