Description of data source: - The Data source is the data scraped from a website ‘investing.com’. We will scrape the data for the top 30 companies and select 8 attributes for analysis. We plan to run it in every 10 mins. Following are the attributes we plan to perform analysis on:- Name, Last, High, Low, Chg, Chg%, vol, time

How to collect data: The data is scraped from a website named ‘investing.com’. We use http requests to hit the website and use beautiful soup to parse the html for us. Using this parsed html we access the data attributes we want and dump this into a table. We then perform ETL on the data and store the data of a Stock into its own table for ease of analysis.

For twitter data, we use the randomized 1% stream to stream out the data and store them in json format into the DB. We then use keyword analysis to filter out tweets related to stock markets.

Measurement and analysis ideas: - We plan to perform sentiment analysis on tweets made on twitter and correlate them to the changes in stock market prices. A future scope of the project would be to predict the change percent of the stocks based on current parameters. We will have to develop a Machine learning algorithm to train a model on this data for the given attributes and then offer insights for the Stock Market

How much data collected: We are scraping 8 attributes for the top 30 companies, which means we will have 30 rows and 8 columns every time we hit the website. The size of the data for 1 cycle is about 32 KB. We plan to run this every 10 minutes, that is we will be generating approximately 4608 KB (4.5 MB) of data per day, totaling it to 32256 KB (31.5 MB) per week.

The stock market refers to the collection of markets and exchanges where regular activities of buying, selling, and issuance of shares of publicly-held companies take place. Such financial activities are conducted through institutionalized formal exchanges or over-the-counter (OTC) marketplaces which operate under a defined set of regulations. There can be multiple stock trading venues in a country or a region which allow transactions in stocks and other forms of securities.

While both terms - stock market and stock exchange - are used interchangeably, the latter term is generally a subset of the former. If one says that she trades in the stock market, it means that she buys and sells shares/equities on one (or more) of the stock exchange(s) that are part of the overall stock market. The leading stock exchanges in the U.S. include the New York Stock Exchange (NYSE), Nasdaq, and the Chicago Board Options Exchange (CBOE). These leading national exchanges, along with several other exchanges operating in the country, form the stock market of the U.S.

Though it is called a stock market or equity market and is primarily known for trading stocks/equities, other financial securities - like exchange traded funds (ETF), corporate bonds and derivatives based on stocks, commodities, currencies, and bonds - are also traded in the stock markets. (For related reading, see "What's the Difference Between the Equity Market and the Stock Market?")