Project 2

Jordan Jeranek

Sam Holland

Stock Data Compilation

We aim to collect weekly stock performance data from the last year and sort into categories for analysis.

X number of stock tickers will be collected with the following performance data:

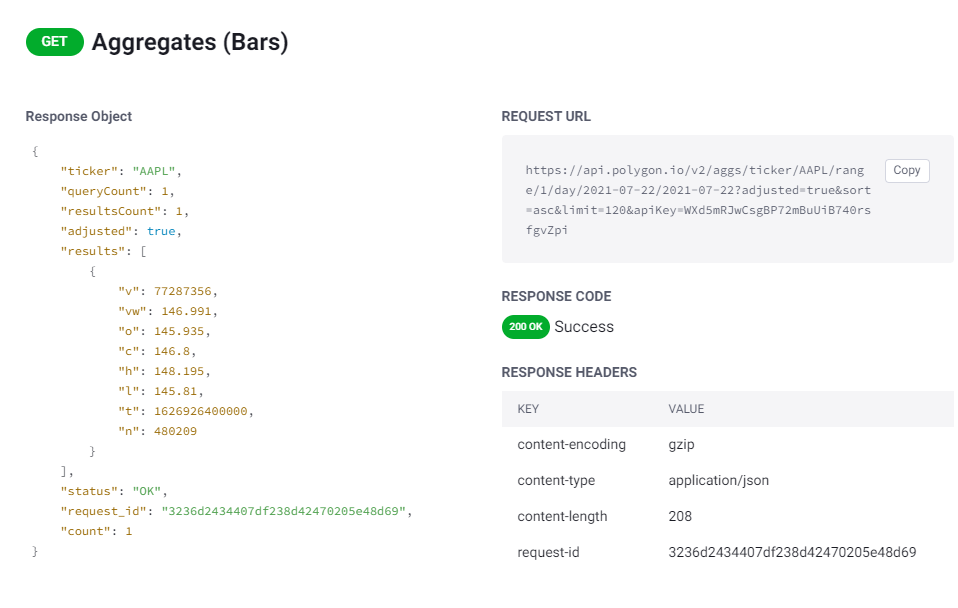
* Opening price
* Closing price
* Weekly high price
* Weekly low price
* Weekly weighted average price
* Weekly trading volume
* Weekly trading volume average

Pull daily data and create weekly data

Stocks can be categorized into the following categories:

* Small/medium/large market caps
* Industries (from GISC MSCI data: [Global Industry Classification Standard - Wikipedia](https://en.wikipedia.org/wiki/Global_Industry_Classification_Standard#:~:text=The%20Global%20Industry%20Classification%20Standard%20%28GICS%29%20is%20an,which%20S%26P%20has%20categorized%20all%20major%20public%20companies.))
* https://stockanalysis.com/stocks/
  + Big tech
  + Pharmecuticals
  + Oil/gas
  + Etc.

Stock data can be pulled from polygon.io API:



Polygon data cleanup:

* Keep only ‘ticker’ and ‘results’ keys from above response
* Rename ‘v’, ‘vw’, etc. to readable key names
* Do we need to load in category data into the same collection as the performance data?

Collections (like tables):

* Stock performance values will be one collection
  + Documents by ticker symbol
* Stock categories (market cap, name, industry) in another collection
  + Documents by ticker symbol
* Top 5 headlines from the year (or something else similar)
  + Title
  + URL
  + Maybe a first paragraph/summary of the article

For Monday:

Sam:

* Performance data
* Web scraping

Jordan:

* NY Times API / headlines / another data source
* Script for loading data into mongo db