Week 8 Yahoo Finance Data

Group 22

April 29, 2021

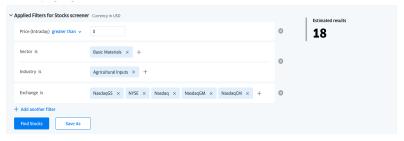
- Obtain Data
 - Select companies via Yahoo Industry Filter
 - Automatically crawl stock prices with Python
- 2 Issues
 - Same query with different results
 - Companies without geocode
 - Profile

2/13

- Obtain Data
 - Select companies via Yahoo Industry Filter
 - Automatically crawl stock prices with Python
- 2 Issues
 - Same query with different results
 - Companies without geocode
 - Profile

Get companies via Yahoo Industry Filter

• Example:



- URL: https://finance.yahoo.com/sector/ms_basic_materials
- copy the table and use regex to get the company symbols
- Industry: agriculture, travel, energy

- Obtain Data
 - Select companies via Yahoo Industry Filter
 - Automatically crawl stock prices with Python
- 2 Issues
 - Same query with different results
 - Companies without geocode
 - Profile

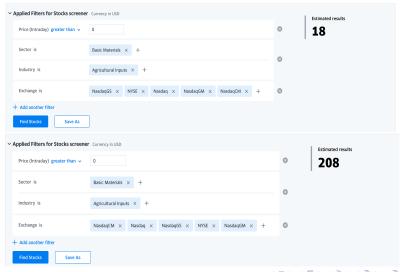
Automatically crawl stock prices with Python

- Input = the *txt* file of one industry
- Output1 = a dictionary {symbol: stock price history, ...}
- Output2 = csv for each company
- History = from 1970-1-1 to 2021-1-1

- Obtain Data
 - Select companies via Yahoo Industry Filter
 - Automatically crawl stock prices with Python
- 2 Issues
 - Same query with different results
 - Companies without geocode
 - Profile

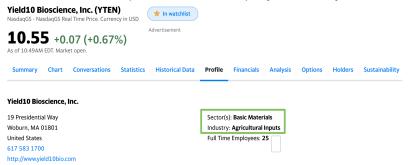
Same query with different results

However, same query may return different table when filtering.



Potential solutions

- try many times, and do intersection
- crawl data from the profile of each company to security its industry



9/13

- Obtain Data
 - Select companies via Yahoo Industry Filter
 - Automatically crawl stock prices with Python
- 2 Issues
 - Same query with different results
 - Companies without geocode
 - Profile

Companies without geocode

- Listed in the U.S., but not located in the U.S.
- Two climate datasets are only about the U.S (wildfire, earthquake)
- Potential solutions:
 - don't care about the geocode
 - use profile like the previous



- Obtain Data
 - Select companies via Yahoo Industry Filter
 - Automatically crawl stock prices with Python
- 2 Issues
 - Same query with different results
 - Companies without geocode
 - Profile

Profile

- Many companies don't have profile page in Yahoo Finance now.
- Can't access all the pages once.
- Potential solutions:
 - use what we have, what we can ensure the industry and location
 - still works on it